공학박사 학위논문

A Semantic Network Analysis as a Method for Understanding Qualitative User Experience in Product Interactions

제품 상호작용 시 정성적 사용자 경험 이해를
위한 의미망 분석 방법

2017년 8월

서울대학교 대학원

산업·조선공학부 인간공학 전공

이 예 림

ABSTRACT

A Semantic Network Analysis as a Method for Understanding Qualitative User Experience in Product Interactions

Ye Lim Rhie

Department of Industrial Engineering

The Graduate School

Seoul National University

Qualitative research provides useful insights with which to analyze the User Experience (UX). This is distinguished from quantitative research by its inductive form of logic and the research aim of understanding holistic phenomena. Since qualitative research aims to identify intangible factors and explore phenomena without simplifying contextual information, it is difficult to exclude a researcher's subjectivity during their analysis. In addition, interpreting and analyzing qualitative materials requires much time and effort. Therefore, this dissertation suggests a systematic research method that utilizes

user expression data to understand UX. The research starts by transforming textual data into numerical representations using semantic network analysis; three major issues were elucidated from the limitations of existing methods: (1) examining the representativeness of the sample size, (2) eliciting important user values (UV), and (3) evaluating product attributes (PA) with numerical inferences.

First, the representativeness of sample size was examined by observing the stability of a semantic network. Among the semantic networks generated from the text, subnetworks were sampled from the original network to vary the sample size. Then, similarities between subnetworks and the original were calculated by applying correlation analysis to node-level centralities. Three case studies that were composed of two interview datasets and one online review data were presented; these proved that this method could be applicable for both small and large samples.

Second, a mixed-method research approach was introduced to suggest appropriate camera shutter press sounds. In qualitative research, important UVs were elicited by analyzing terms with high centralities in a semantic network. The elicited UVs were then used as questionnaire items in quantitative research to represent UV with numerical values. The result demonstrated user satisfaction models for shutter press sounds and the relationships between UV and PA by adopting the concept of psychoacoustic variables.

Third, the importance of UV and their relations to PA were examined based on qualitative research on vacuum cleaners. Seven types of network centrality were used to weight the UVs, which resulted in UX quantification models. These models' goodness-of-fit were compared to the results of quantitative research. Then, the links between UV and PA nodes were identified. Since

statistical analysis without a proper theoretical interpretation may mislead users, qualitative data can assist quantitative research by examining the sematic associations between UV and PA.

Compared to traditional qualitative studies, the proposed method in this dissertation has a competitive edge for reducing the cost, effort, and subjectivity. Determining the smallest sample size that can achieve network stability is a novel data collection strategy that attempts to maximize effectiveness while minimizing both cost and effort. Utilizing this method allows UX researchers and practitioners to collect the optimal sample size by gradually increasing their sample sizes. Important UVs were elicited in the process of evaluating UX, and their importance was quantified to build a UX quantification model. Transferring qualitative descriptions to the quantitative models allows researchers to understand UX more efficiently by reducing the process of collecting numerical data on each UV. Lastly, important PA and their relations to UV were identified. Although centrality measures were not proportional to the correlation level, semantic associations between UV and PA could be identified. Considering that huge amounts of text data are being generated and collected every day, the suggested method is expected to be useful for practical applications when developing products.

Keywords: User Experience, Qualitative research, Semantic network analysis, Network stability, Data representativeness, User value, Product attribute

Student Number: 2012-21066

CONTENTS

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1. Background and motivation

Because customer satisfaction directly affects success of a product, it is vital to understand user experience (UX) in developing a product (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009). However, there is no universal method to evaluate UX because of its dynamic, context-dependent, and subjective nature. Hassenzahl and Tractinsky (2006) defined UX with three dimensions of "user's state," "system properties," and "context-of-use," whose characteristics vary with user groups and use cases. The ISO standard also emphasized the individual and personal characteristics of UX, by defining "*a person's perceptions and responses that result from the use or anticipated use of a product, system, or services* (ISO 9241-210, 2009)".

To understand UX, researchers have identified UX dimensions, surveyed quantitative scores on UX dimensions, and conducted a statistical analysis on the collected numbers (Edwards & Bagozzi, 2000; Willett & Keiley, 2000). In process of identifying UX dimensions, researchers conducted qualitative researches such as observation, participation, interviewing, and ethnography, to get a better understanding of the subject matter (Kujala, Roto, Väänänen-

Vainio-Mattila, Karapanos, & Sinnelä, 2011; Marshall, 1996; Park, Han, Kim, Moon, & Park, 2014; Patton, 2005; Roto, Rantavuo, & Väänänen-Vainio-Mattila, 2009; Vermeeren et al., 2010; Visser, van Biljon, & Herselman, 2013). However, in these days, it became necessary to analyze user expression data more efficiently. As a result of the proliferation of Internet and Communication Technologies (ICT), numerous customers voluntarily express their feelings and preferences related to the product or services. Actually, Seth (2008) reported that 80% of the data are unstructured, and mainly in a text type. Therefore, interpreting qualitative data and extracting meaningful information have become key tasks to researchers (Morstatter, Pfeffer, Liu, & Carley, 2013).

Unstructured data collected from qualitative research can provide useful insights that structured data cannot provide. However, it cannot be analyzed by statistical testing, since qualitative research is distinguished from quantitative research by its inductive form of logic, and its research aim to understand the holistic phenomenon. Whereas quantitative research generalizes its results by applying statistical inference, qualitative research emphasizes identifying intangible factors and exploring the phenomena without simplifying the contextual information (Marshall, 1996; Patton, 2005). Therefore, researchers can hardly exclude their explicit knowledge and qualitative interpretation in analyzing qualitative data, as Ngulube (2015) stated, "*The major challenge in discussing the interpretation of qualitative data stems from the fact that interpretation is regarded as an art that is not amenable to formal rules*". Creswell (2012) also emphasized the subjectivity of interpreting qualitative data as he stated, "*Data analysis is not off-the shelf; rather it is custom-built, revised and choreographed*".

To overcome such drawbacks, there has been attempts to quantify UX. Interview techniques such as repertory grids (Kelly, 1955), Q-sort, and laddering (Reynolds & Gutman, 1988) were suggested to quantitatively evaluate text data, by collecting numerical information in addition to the verbally expressed data. However, these techniques required a specific data format or a predetermined structure, which inevitably involved researchers' subjective judgment.

In the meantime, there are methodologies that take a grounded theory approach, not to involve researchers' subjective bias. Among, content analysis has been prevalently applied to qualitative researchers. It focuses on extracting meaningful information from various types of sources including text, image, and video (Downe-Wamboldt, 1992). It takes one step further from merely describing the current status (Morgan, 1993), by inferring the importance of concepts from the frequency of words and instances (Kondracki, Wellman, & Amundson, 2002). However, the method has a limitation in identifying relationships between concepts.

For large corpus of text, Latent Semantic Analysis (LSA) has been suggested to organize text-object matrix into smaller dimensions using a singular-vector decomposition (Foltz, 1996; Landauer, Foltz, & Laham, 1998). However, since the power of LSA comes from a large volume of training corpora, the methodology can only be adopted for a collection of documents and online resources (Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988; Kulkarni, Apte, & Evangelopoulos, 2014; Müller, Schmiedel, Gorbacheva, & vom Brocke, 2016).

By applying data mining techniques on customer reviews and ratings, satisfaction level of customers can be predicted from text data (Archak, Ghose,

& Ipeirotis, 2007; Fang & Zhan, 2015; Malandrakis, Potamianos, Iosif, & Narayanan, 2013; Socher et al., 2013). The sentiment analysis infers subjective feelings of users (i.e. positive, negative, and neutral) from text data (Tsytsarau & Palpanas, 2012; Van Atteveldt, 2008), which enables identifying product properties that give positve or negative influences. However, the studies usually focus on classifying users' consequent feelings in a summative approach, shading little attention in eliciting user values (UV) and product attributes (PA) that affect customer purchase.

Semantic network analysis counts relationships between terms which are collected from natural language. It represents a semantic association between concepts, and observes structural relations and patterns between a given set of objects (Borgatti, Mehra, Brass, & Labianca, 2009; Freeman, 2004). Van Atteveldt (2008) points out that semantic network analysis has an advantage over content analysis in overcoming an abstraction gap, which refers to the fact that content analysis represents the collected data into higher-level concepts. Depend on research purpose, the terms that appeared in the same sentence, paragraph, or answer are assumed to have relationships, and defined to "co-occur". Kim, Lim, and Yun (2016) applied the method to figure out the motivations underlying teenagers' Internet use, and proved its effectiveness comparing to the laddering technique. Meanwhile, Morstatter et al. (2013) verified the effect of sample size by comparing data retrieved using Twitter's Streaming API and Firehose. Although there have been several studies reporting the effectiveness of network analysis (Callon, Courtial, Turner, & Bauin, 1983; Chomsky, 1980; Hoser, Hotho, Jäschke, Schmitz, & Stumme, 2006; Kim et al., 2016; Morstatter et al., 2013), no study has suggested a systematic research method of using network analysis in understanding UX.

The problems of using user expressions in understanding UX are elucidated with three issues, following the steps of data collection, interpretation, and product design.

First, while collecting qualitative data, previous qualitative researchers suggested appropriate sample size depend on their subjective judgement. Since researchers cannot investigate whole user groups due to time and budget constraints, sampling strategies are generally required; initially, researchers group categories of participants and proceed to “purposive” or “selective” sampling. In these non-probabilistic methods, researchers proposed appropriate sample sizes based on their experiences. For example, Morse (1994) proposed investigating at least six participants for phenomenological studies, 40–60 for ethnographic and grounded theories, and 100–200 for qualitative ethological studies. Baum (2003) considered 12–20 participants were sufficient for maximum variation sampling (Glaser, Strauss, and Strutzel, 1968). Crouch and McKenzie (2006) stated that 20 participants are sufficient for an in-depth inquiry. Although it is important to ensure the representativeness of data in terms of reliability, only few studies have succeeded in judging the representativeness without involving a subjective bias.

Second, it is hard to exclude experts’ explicit knowledge and qualitative interpretation in identifying UX dimensions, as it is highly subjective and context-dependent. There are some arguments that UX cannot be reduced to numbers, but it would be still meaningful if it helps comparing and predicting the quality of UX (Law & van Schaik, 2010). Consequently, researchers have identified UV, and decided how to quantify them. Rather than conducting an additional experiment to obtain quantitative inferences, it would be much more practical if we can quantify the importance based on text data. To solve this problem, a semantic network analysis can be used, which calculates the influence of a node in a network by calculating directly and indirectly linked nodes. However, up to my knowledge, no study has applied a semantic network analysis to quantitatively analyze UX.

Third, the attempt to identify relationships between UV and PA achieved only marginal success. Quality Function Deployment (QFD) provides a framework to illustrate the importance of each PA (Cohen, 1988; Hauser & Clausing, 1988), Kansei engineering provides a theory to match user perception against the physical product attributes (Nagamachi, 1995), and UNISON framework integrates subjective and objective factors of UV to relate PA (Chien, Wang, & Wang, 2007). However, these methods still require expert's explicit knowledge in extracting PA of a product. In this paper, important PAs and their relations to UV were identified based on user expression data.

1.2. Research objective

The purpose of this paper is to provide a systematic method to reduce researchers' subjectivity in understanding UX. Three problems of existing qualitative researches were clarified by: (1) examine representativeness of sample size, (2) identify UV, and (3) relate UV and PA. The research starts with generating a semantic network, which enables the numerical representation of keywords mentioned by users.

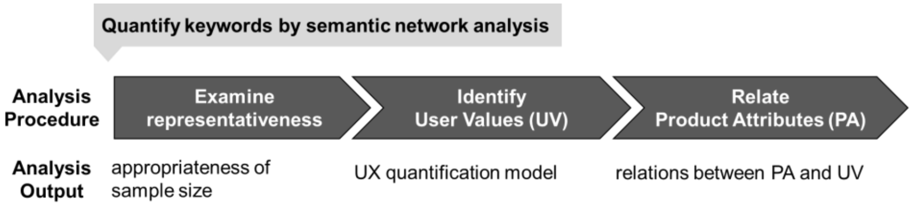


Figure 1.1. Purpose of dissertation

1.2.1. Examine representativeness

The first part of this paper suggests a method that quantitatively assesses the stability of textual data. Stability means the degree of invariant or unchanging content regardless of the new coming information, thereby indicates the representativeness of data. Whereas the problem of representativeness is an important issue which directly affects the quality of data, it has been subjectively determined by UX researchers. Or, theoretical sampling method, which requires too much time and effort, was applied. Therefore, this paper proposes a novel approach to measuring the stability of textual data by adopting semantic network analysis.

1.2.2. Identify user values (UV)

Due to its dynamic, context-dependent, and subjective characteristics, UX researchers have difficulties in defining UX factors, in other words, UV. Therefore, contrary to the quantitative study which aims to generalize the results using statistical inference, UX researchers focus on identifying descriptive factors to understand contextual information. Details or scenarios can be a good method in understanding UX, but it is also necessary to elicit important UVs that determine user's satisfaction. In this dissertation, important

UVs were elicited based on network centrality measures, and were used as questionnaire items for quantitative study in a mixed-method research. Going one step further, the numerical values of network centralities were also used to give weights to UVs in building UX quantification models.

1.2.3. Relate product attributes (PA)

UX researchers make a number of decisions while developing products or services. In order to make a better decision, they should be aware of the relationships between UV and PA. However, existing methods usually require a predefined hierarchical structure or product property, which require experts' explicit knowledge or additional surveys. There are two types of PA: externally visible and measurable features, and those of which are difficult to be observed or measured without a domain knowledge (Sharp, Rogers, & Preece, 2007). For the former type of PA, user expression data can be utilized in identifying important PAs in relation to UVs. Since a network analysis examines semantic level of association between concepts, the method will help avoiding spurious correlation, which describes a relationship between two variables that are correlated, but happens just by chance or due to an unseen third variable.

1.3. Dissertation outline

The structure of this dissertation is shown in Figure 1.2. The main topic is to suggest a systematic research method of using qualitative text data. Following steps of examining data representativeness, identifying UV, and relating PA to UV, the methods of utilizing qualitative text data are introduced.

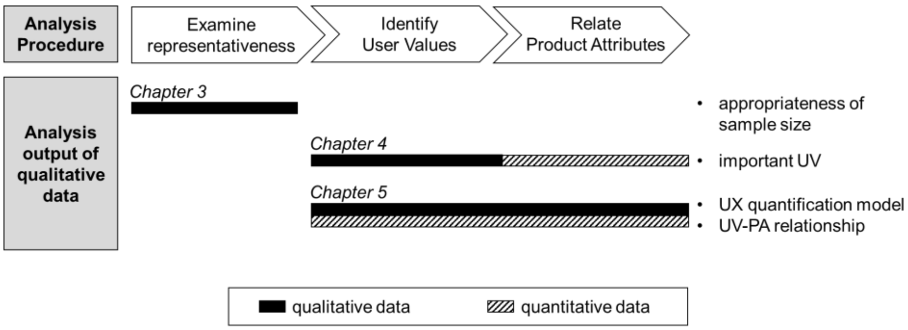


Figure 1.2. Dissertation outline

Chapter 2 reviews related work. A brief definition of semantic network analysis, measures of network centrality, and the concept of network stability are introduced. This chapter also examines the existing researches in aspect of sample size, UX evaluation techniques, and product design.

Chapter 3 introduces a method to measure the stability of textual data by adopting semantic network analysis, with case studies on two in-depth interview transcripts and one online customer review data. Among the semantic networks generated from text, subnetworks are sampled from the original networks until the representativeness of each sample size is determined. Then, similarities between the subnetworks and the original network are calculated by applying a correlation analysis to determine whether stability occurs.

Chapter 4 applies a network analysis in a mixed methods research. From qualitative research, top 10 keywords with the highest degree, closeness, betweenness, and eigenvector centralities were analyzed to elicit important UVs. Then, these UVs were used as questionnaire items for a consequent quantitative research. From this study, user satisfaction models for two user groups were suggested for a camera shutter press sound.

Chapter 5 utilizes the value of network centralities to build UX quantification model. This chapter introduces two issues of (1) suggesting UX quantification model, and (2) identifying relationship between UV and PA based on qualitative data. First, important UVs on vacuum cleaner were elicited, and UX quantification models were proposed by utilizing centrality measures as UV weights. The goodness-of-fit of these models were compared to the result of quantitative study. Second, the relationship between UV and PA were identified based on a simple hypothesis: if UV and PA are related, terms will co-occur. We generated a subnetwork for each PA, and calculated UV centralities on each subnetwork. The result was used to assist the result of correlation analysis. Since network analysis reveals semantic level of association between two concepts, the method will be able to assist quantitative research in avoiding spurious correlation.

Chapter 6 discusses findings obtained from this study. This chapter also describes limitations and implications of this dissertation.

CHAPTER 2

Literature Review

2.1. Semantic network analysis

2.1.1. Definition

Semantic network analysis is defined as *"a study in which word associations in texts are analyzed, and those word associations represent the meaning of inherent to the data*" (Doerfel, 1998). Semantic network analysis applies social network analysis metrics, representing nodes as concepts and an edge as a link between two concepts (Carley, Columbus, & Azoulay, 2012; Hoser et al., 2006; Tanenbaum & Brand, 2008). The term is referred by different names depending on the granularity of the concepts, or whether they focus on nouns or verbs (Diesner & Carley, 2004). Types of semantic networks include concept maps (Carley, 1993), knowledge networks (Popping, 2003), mental models (Carley, 1997a), meta-networks (Diesner & Carley, 2008), network text analysis (Carley, 1997b), and relational content analysis (Van Atteveldt, 2008). Throughout this paper, "semantic network analysis" is used as a general term indicating a technique that uses mathematical algorithms to represent the concepts and their semantic relationships.

A semantic network represents relations between concepts. Concepts can be actors, issues, abstract values, and general words. The associations between concepts are often determined by their co-occurrence. This approach of extracting associations is not a new, as it has long been used in the field of contents analysis (Harway & Iker, 1969). By using a semantic network, previous studies have proved semantic distances between two concepts could be measured (Rips, Shoben, & Smith, 1973; Rodríguez & Egenhofer, 2003). In other words, semantic distances between two concepts are calculated by generating semantic networks.

Initial model of semantic structures had hierarchical structure (Collins & Quillian, 1972; Quillian, 1969) and simple pairwise connections (Estes, 1982). Quillian (1969) first defined the concept of nodes and links to represent concepts and relations between them. Based on this hypothesis, Wilkins (1971) highlighted the concept of "semantic distance," semantically measuring "similarity" of words within a network. Several other researchers revealed linguistic or textual influences by comparing the characteristics of network structures (Callon et al., 1983; Chomsky, 1980).

2.1.2. Co-occurrence

Co-word analysis has been widely used by researchers who tried to measure the semantic similarities in text-mining field (Feng, Zhang, & Zhang, 2017; Iosif & Potamianos, 2015; Lund & Burgess, 1996). The basic idea is that words with similar meanings will tend to occur in similar contexts, and hence word co-occurrence statistics can provide a natural basis for semantic representations (Bullinaria & Levy, 2007). The words are represented with vectors based on distance metrics such as Euclidean distance, then statistical analysis such as

Multidimensional scaling (MDS) and Principal Component Analysis (PCA) is applied. This approach has been profoundly used by several researchers. For example, LSA technique (Foltz, 1996), which uses a singular vector decomposition (SVD) method on co-occurrence matrix, has been widely used to extract latent dimensions of data.

A link of a semantic network is generated based on a co-occurrence network. Co-occurrence matrix is obtained by counting frequencies of a pairwise term appeared in a unit (i.e. sentence, paragraph, document) called "word boundary". The word boundary is defined by research objective; if a researcher selects paragraph as word boundary, each word that appears in a paragraph gets a fully connected network for each paragraph. Let terms *a*, *b*, *c*, *d* and *e* appear in the same word boundary. Then the co-occurrence matrix becomes 5×5 symmetric matrix as shown in Figure 2.1.

	a	*b*	*c*	*d*	*e*
a	-	30	26	19	18
b	30	-	5	50	6
c	26	5	-	4	27
d	19	50	4	-	3
e	18	6	27	3	-

Figure 2.1. Example of co-occurrence matrix

2.1.3. Network statistics

Networks are composed of nodes and links, and are generally described as a graph $G = (V, E)$, where $v \in V$ is a node (vertex) and $e \in E$ is an edge. Each node represents a concept, and the edges represent relationships between two concepts, called the co-occurrence. The statistics of G (i.e., network-level statistics) explain the shape and cohesion of a network (Borgatti et al., 2009), whereas the statistics of v explain the network centralities (Freeman, 1978).

Measures of the network shape (average distance, diameter, breadth, and compactness) represent its descriptive characteristics. Cohesion measures calculate characteristics for regions within the network, and typically include the average degree, density, connectivity, and clustering coefficient. The density of a network is defined as the proportion of linked networks to the number of possible edges, and connectivity calculates the number of nodes that must be removed to completely isolate the region. The clustering coefficient is the probability that two randomly chosen nodes are neighbors, which reflects the number of connected nodes in each cluster (Watts & Strogatz, 1998). If $k(i)$ represents the number of neighbors of node i and B is the total number of edges between all nodes that are related to node i, the number of connected nodes is $k(i)(k(i)-1)$. Thus, the local clustering coefficient of node i is $C_i = B/[k(i)(k(i)-1)]$, and the global clustering coefficient is $\overline{C} = V^{-1}\sum_{i=1}^{V} C_i$. In the context of semantic network analysis, cohesion statistics represent the overall similarity between two concepts, and can be used to examine the text quality. Some researchers found that text quality decreases as the clustering coefficient increases (Antiqueira, Nunes, Oliveira Jr, & Costa, 2007; Martin, Pfeffer, & Carley, 2013; Pardo, Antiqueira, Nunes, Oliveira, & Costa, 2006).

The network-level characteristics described above represent the general properties of texts. For example, small-world network structures generally have short average distances and diameters, that associated words with a significant fraction in neighboring clusters are considered to be related (Steyvers & Tenenbaum, 2005). On the other hand, node-level centralities represent statistical features of each node. At the node level, the most frequently used measures are degree, betweenness, eigenvector centralities, and closeness, which indicate the role and importance of each node. High degree centralities imply that many neighboring nodes are directly mentioned, whereas high betweenness centralities represent the location of nodes in a global sense. Eigenvector centralities illustrate the overall structure of a network, and quantify the differences between ontologies (Hoser et al., 2006), and closeness represents the number of paths connecting two other nodes that a designated node is connected to. Centralities for bidirectional networks are described below.

2.1.3.1. Degree centrality

Degree centrality can be defined as the number of ties incident upon a node. The centrality focuses on local position in the network, therefore cannot incorporate indirect connections.

Degree centrality is defined as the number of directly linked nodes. The normalized degree centrality is calculated by,

$$C_D(v_i) = \deg(v_i)/(n-1) \tag{1.1}$$

where $\deg(v_i)$ is the number of connected nodes (Wasserman & Faust, 1994).

Bonacich power is a modified version of degree centrality that takes similar perspective with eigenvector centrality; centrality of a vertex can be represented by the summation of the centrality of the vertices it is connected to. Given an adjacency matrix A, the centrality of vertex i, $C_P(v_i)$, is calculated by

$$C_P(v_i) = \Sigma A_{ij} (\alpha + \beta C_j) \quad (1.2)$$

where α and β are parameters; UCINET software (Borgatti, Everett, & Freeman, 2002) selects the largest permissible value for α, so that the square root of the sum squares of the vertex centralities becomes the size of the network. β implies the attenuation factor. Zero is proportional to the number of connected nodes, positive values give weight to being connected to powerful vertices, and negative values give weight to being connected to low powered vertices.

2.1.3.2. Betweenness centrality

Beweennness centrality is the most widely used path-based measure. The measure focuses less attention on access to information for a powerful node, but rather concentrates on the power resulting from being on the shortest path among others. Borgatti et al. (2002) described betweenness centrality as “the number of times that any actor needs a given actor to most efficiently reach any other actor”.

Freeman betweenness centrality is the most widely used method of path-based measures that present the importance of a node works as a brokerage, calculated by

$$C_B(v_i) = \Sigma_i \Sigma_j g_{jik} / g_{jk} \; (i \neq j \neq k) \quad (1.3)$$

where g_{kij} is the number of geodesic (shortest) paths from vertices v_j to v_k , and

g_{jik} is the number of geodesics that pass through node i. Although there are many possible paths between two nodes, Freeman betweenness centrality only considers the shortest path connecting two nodes. The equation represents the location of a node in the graph and illustrates the role of an intermediary (Freeman, 1978). Normalized value divides $C_B(v_i)$ by the number of pairs of nodes except v_i, which is calculated by $(n - 1)(n - 2)/2$.

Flow betweenness centrality is different from Freeman betweenness centrality, as it assumes that two nodes will not always use the shortest path between them. Thus, the measure considers all pathways that go through v_i and v_j. Let m_{jk} indicates the amount of flow between v_j and v_k ($j < k$), which must pass through v_i for any possible flows. Then flow betweenness of v_i is calculated by the sum of all m_{jk} (Freeman, Borgatti, & White, 1991), as represented with following equation,

$$C_F(v_i)=\Sigma_i \Sigma_j m_{jik} / m_{jk} \ (j < k;\ i \neq j \neq k) \tag{1.4}$$

2.1.3.3. Eigenvector centrality

Eigenvector centrality indicates the importance of a node in the entire network. It is defined as the principal eigenvector of the adjacency matrix defining the network (Borgatti et al., 2009). The eigenvector centrality for node v_i linked with node v_j is

$$C_e(v_i) = \lambda^{-1} \Sigma_j A_{ij} C_e(v_j) \tag{1.5}$$

where $\lambda C_e= A^T C_e$ and λ is the corresponding eigenvalue. The equation can be interpreted that higher eigenvalue λ indicates higher influence of the nodes. The normalized eigenvector centrality divides $C_e(v_i)$ by the maximum difference possible C_e^{max}.

2.1.3.4. Closeness centrality

Closeness centrality is a distance-based measure that assumes relative differences in the length of the paths between nodes constitute an important factor in determining an actor's centrality and power.

Freeman closeness is the best-known method to calculate closeness centrality. It calculates the geodesic distance of a vertex and the other nodes in the entire network. The closeness of v_i refers to the sum of geodesic distances from node v_i to all n-1 others in the network. For disconnected nodes, a default distance was assigned as total number of nodes in a network plus one. The formula is represented as

$$C_{clo}(v_i) = [\Sigma^{N}_{j=1}\, d\,(i, j)]^{-1} \qquad (1.6)$$

based on the length of the average shortest path between a vertex and all vertices in the graph (Freeman, 1978).

Average Reciprocal Distance (*ARD*) is another measure of closeness centrality, which is the reciprocal of the sum of the length to all other nodes. Therefore, higher value of ARD means the greater connectedness of the node. The average of reciprocal distance is represented as below (Borgatti, 2006).

$$C_{ARD}\,(v_i) = \Sigma^{n}_{j=1}\, d(i, j)^{-1} / \,(n - 1) \qquad (1.7)$$

Contrary to the Freedman's closeness centrality, whose distances are summed first before being placed in the denominator, ARD is summed first and summed.

2.2. Sample size

2.2.1. Reliability of qualitative text data

The reliability of text data is defined by stability, reproducibility, and accuracy (Krippendorff, 2004). Stability refers to the invariance of content regardless of new information and thereby indicates the representativeness of data, reproducibility examines the similarity of judgments between different coders, and accuracy implies the extent to which a process meets a particular standard. Qualitative researchers have mainly addressed the question of reliability in terms of reproducibility, which matters when natural language is coded into a semi-structured or structured format (Campbell, Quincy, Osserman, & Pedersen, 2013; Hruschka et al., 2004; Kurasaki, 2000; Yu, Jannasch-Pennell, & DiGangi, 2011). Accuracy generally means finding typos and errors based on spelling standard, but the accuracy in content analysis implies how much coders followed the standard, in which researchers already have established (Krippendorff, 2004). Therefore, the verification of accuracy is only required when there is a standard to observe. Most studies have suggested quantitative criteria of reproducibility and accuracy by calculating the number of agreements and disagreements between coders (Campbell et al., 2013; Miles & Huberman, 1994).

However, almost no studies have numerically calculated the stability of qualitative data. Instead, several rule-of-thumb guidelines have emerged: Morse (1994) proposed investigating at least six participants for phenomenological studies, 40–60 for ethnographic and grounded theory studies, and 100–200 for qualitative ethological studies. Crouch and McKenzie (2006) stated that 20 participants are sufficient for in-depth inquiries. Meanwhile, the concept of "theoretical sampling" (Glaser et al., 1968) was also proposed. It takes an inductive approach based on grounded theory to iteratively collect and categorize concepts until no new categories appear, which is called "theoretical saturation." However, this method requires too much time to examine a large volume of data, because researchers must repeatedly collect, investigate, and analyze the data. Although there have been studies to automate the classification of categories and calculate reliability, the methods are only effective on large amounts of data, and involve subjective bias in defining categories and similarity of concepts (Bratko & Šuc, 2003; Chang, Lin, & Wang, 2009; Ngai, Xiu, & Chau, 2009).

2.2.2. Sample size of HCI studies

In the field of Human–Computer Interaction (HCI), it is a central issue to clarify if problems were sufficiently identified. Therefore, discussions on sample size arose in identifying usability problems. To address the reliability of usability studies, Virzi (1992), Wright and Monk (1991), and Nielsen (1994) have presented mathematical model of problem discovery rates. They showed that it was possible to detect most of usability problems with the first three to five participants.

Wright and Monk (1991) suggested an equation, D=1-(1-p)^n, to demonstrate the relationship between the problem discovery rate across subjects (p), number of subjects (n), and percentage of events to be discovered (D). Similarly, Nielsen and Landauer (1993) used Poisson model to describe the number of usability problems found with the equation, Found (i)=N(1-(1-λ)^i). Here, λ is the probability of detecting a problem, N is the total number of problems, and i is the number of users. Utilizing this equation, Nielsen (1994) reported that five subjects are sufficient to solve 77–85% of usability problems with mean problem discovery rate of 0.28–0.30 when using the think-aloud method.

In a line of research, Lewis (2001) presented that the proportion of discovered problems were overestimated when sample size is small. He suggested compensating the overestimated estimate (p_{est}) by using Good-Turing estimation discounting method and normalization method. The resulting adjustment is presented as, $P_{adj} = 1/2\ [(p_{\mathrm{est}} - 1/n)\ (1 - 1/n)] + 1/2[p_{\mathrm{est}} / (1+GT_{adj})]$, where GT_{adj} is the Good-Turing adjustment which is calculated by the proportion of the number of problems occurred once divided by the total number of different problems. Other researchers also explored the effect of experimental condition on sample size, and emphasized revealed the risks of using only five participants (Faulkner, 2003; Hwang & Salvendy, 2010; Law & Hvannberg, 2004; Savoy, Guo, & Salvendy, 2009; Schmettow, 2012).

However, the suggested equations may not be appropriate for UX studies, as UX issues cannot be counted or discovered as usability issues were found. Since UX researchers adopt a more holistic approach that encompasses the objective approach of usability studies (Hassenzahl & Tractinsky, 2006; Lallemand, Gronier, & Koenig, 2015), a new approach to measure the appropriateness of sample size is required.

2.3. User experience (UX) evaluation techniques

2.3.1. User value (UV)

User value (UV) is a key factor of UX that makes users purchase the products (Kim et al., 2016; Park, Han, Kim, Oh, & Moon, 2013). It is defined as a psychological dimension of UX (Kujala & Väänänen-Vainio-Mattila, 2009) that results from the interaction between products (Boztepe, 2007). Park and Han (2013) also defined user values as "desirable states of existence or modes of behavior which are satisfied when using a certain product or service". Rokeach (1968) suggested eighteen terminal values such as true friendship, mature love, and self-respect, with eighteen instrumental values, for example, cheerfulness, ambition, and forgiveness. Holbrook (1999) framed user value with three dimensions, which are extrinsic versus intrinsic, self-oriented versus other-oriented, and active versus reactive. From the suggested framework, user value can be categorized with "utility value," "social significance value," "emotional value," and "spiritual value (Boztepe, 2007)". Utility value focuses on the efficient aspect of products that make users achieve goals, whereas social significance value mainly considers socially oriented benefits such as making relationship with others, or attaining ownership of a product. Emotional value refers aesthetics and fun elements, and spiritual value represents supernatural beliefs.

Several techniques, such as Values and Lifestyles (VALS) questionnaire (Woodruff, 1997) and interview based on means-end model (Zeithaml, 1988), can be generally applied to measure the user value. With increasing needs to minimize subjective bias in identifying user value, non-hierarchical network analysis has been applied on interview transcripts (Kim et al., 2016). Kim et al.

(2016) adopted semantic network analysis to assess the motivations underlying teenagers' Internet use, and proved its effectiveness comparing to the laddering technique.

2.3.2. Quantification model

"UX model" is an abstract representation or an approximation of an underlying UX theory (Law, van Schaik, & Roto, 2014). Through quantification, researchers can attain a certain level of meaningfulness and validity. Thüring and Mahlke (2007) suggested Components of User Experience (CUE) model which divides instrumental quality (i.e. controllability, effectiveness, and learnability), non-instrumental quality (i.e. visual aesthetics, haptic quality, symbolic), and emotional reactions (i.e. subjective feelings, motor expressions, physiological reactions). Based on CUE model, Law et al. (2014) conducted a survey – UX Measurement Attitudes Survey (UXMAS) – and classified measurable and non-measurable dimensions (Kujala et al., 2011). Although it is still disputable if user values are quantifiable, many researchers have made efforts to measure UX. Desmet (2005) developed "PrEmo" which visually assesses product-related emotions with animated cartoon characters. By using PrEmo, users reported their emotional responses by selecting an emoticon that corresponds with their feelings. Kujala et al. (2011) proposed "UX curve" which enables users represent their experience on product quality by a line chart on time axis, and Tuch, Bargas-Avila, Opwis, and Wilhelm (2009) assessed psychophysiological responses such as reaction time and electrodermal activity of users.

However, most of researchers measure UX by using a questionnaire-based survey. Various questionnaire items were developed depend on research objective and product/service properties. For example, “AttrakDiff2” (Hassenzahl, Burmester, & Koller, 2003) is a questionnaire composed of 28 items which measures pragmatic quality, hedonic stimulation, hedonic identification, and attractiveness. The method has been applied in designing skins of MP3 player (Hassenzahl, 2005), task-oriented software (Isleifsdottir & Larusdottir, 2008), and a mobile application (Van Dantzig, Geleijnse, & van Halteren, 2013). For more specific context-of-use, researchers first investigated UX dimensions through qualitative studies, and developed questionnaire items to evaluate. For example, smart products were characterized with seven variables of autonomy, adaptability, reactivity, multi-functionality, ability to cooperate, humanlike interaction, and personality (Rijsdijk & Hultink, 2009). Additionally, UX on smartphones was investigated with five variables of comfort, relationship, convenience, beauty, and social status (Park & Han, 2013).

Based on the assumption that UX elements construct UX exclusively (Kim, Han, Park, & Park, 2015), researchers have applied linear model to the elements of UX to evaluate the satisfaction level. More specifically, multiple regression model (Asche & Kreis, 2014), structural equation model (Hassenzahl, Diefenbach, & Göritz, 2010; Knijnenburg, Willemsen, Gantner, Soncu, & Newell, 2012; Park et al., 2014), and nonlinear models were applied. Park et al. (2013) compared simple linear model, polynomial model, S-shaped value model, conjunctive model, and disjunctive model, and revealed all models were valuable except for disjunctive model.

2.4. Product design

2.4.1. User Centered Design (UCD)

Incorporating user's various demands into user interface increases user satisfaction, and reduces design cost (Chaffin & Nelson, 2001). ISO 9241-210 (2009) proposed UCD process for product developers, which is composed of four iterative steps: understand and specify the context of use, specify the user requirements, produce design solutions, and evaluate designs against requirements. If the solution does not meet user requirements, designers should proceed these steps again.

With a comprehensive understanding of users, a researcher becomes able to find out whether the product can fulfill customer's expectations or not. Therefore, UCD process first recommends to understand the context of using products. Several research methods of context of use analysis, field study, survey, observation, and diary keeping methods are used (Bruseberg & McDonagh-Philp, 2001; Maguire, 2001). After defining a context of use, user requirements should be established. Qualitative analysis including in-depth interview, scenario method, and FGI (Focus Group Interview) are carried out. This step includes mapping task and function of a product (Bruseberg & McDonagh-Philp, 2001). Delivering a design solution requires an innovative creativity. In this step, ideation techniques such as brainstorming, parallel design, and TRIZ (Altshuller, 1984; Wang, Chen, Lin, & Wang, 2005) or design guidelines and standards (Bevan, 2001; Nielsen & Molich, 1990) are utilized. When the outcome of this solution is realized as a prototype, evaluation techniques are applied. According to the research purpose, a prototype from low to high fidelity is developed. With a technological development, prototyping become easier and faster by using 3D printing techniques or simulation environments (Smith & Dunckley, 2002).

2.4.2. Design method

Product developers make a number of decisions in designing a product. To achieve better outputs, researchers should collect user requirements and map these demands to the product attributes or properties.

One of the most popular methods is Quality Function Deployment (QFD), which relates user requirements with the product attributes using House of Quality (HoQ) matrix (Cohen, 1988; Hauser & Clausing, 1988). During process, product developers determine the association strength between attributes and user values. They use their specialized knowledge, or survey customers to identify the relatedness between product attributes and user requirements (Moghimi, Jusan, Izadpanahi, & Mahdinejad, 2017; Myint, 2003; Schauerte, 2013).

Interpretive Structural Modeling (ISM) is a method that helps analyzing the hierarchical relationship between components, especially for complex systems. This method allows researchers to relate PA to an issue or a problem systematically (Warfield, 1973, 1978). To define relations between PA and UV, the method recommends to collect experts' opinions based on management techniques including brain-storming (Attri, Grover, Dev, & Kumar, 2013).

Whereas the above methods directly evaluate connectedness between user requirements and PA, evaluative approach indirectly infers the relationship by applying statistical techniques as described below.

Design of experiment (DoE) is one of the most traditional method to evaluate the best combination of PA. The method evaluates the effects of multiple factors on the performance, based on a statistical background. However, DoE approach requires much time and efforts in developing samples and conducting experiments.

Conjoint analysis is one of the most widely used methods in the field of marketing (Green & Srinivasan, 1978). It basically surveys customer reactions compared to the alternatives, and statistically calculates the effects of PAs. It evaluates the relative importance of each PA (i.e. color, size), and utility of properties (i.e. red, blue, green).

Kansei engineering is a technology that links physical product property, perception, cognition, and emotion (Nagamachi, 1995). It collects customer's Kansei experience and establishes mathematical prediction models by classifying physiological and psychological aspects of a human (Nagamachi, 2002).

CHAPTER 3

Evaluating Representativeness of Unstructured Text Data

3.1. Overview

To understand UX, researchers usually have focused on identifying descriptive factors to encompass contextual information (Kujala et al., 2011; Kujala, Vogel, Pohlmeyer, & Obrist, 2013; Park et al., 2013). They collect customer opinions by a naturalistic approach, in order to explore UX dimensions and their implications (Marshall, 1996; Patton, 2005; Visser et al., 2013). However, little attention was given on the aspect of data reliability, although it can mislead researchers in understanding users (LeCompte & Goetz, 1982; Ramsey & Hewitt, 2005).

The studies of network theory (e.g., (Albert, Jeong, & Barabási, 1999; Dorogovtsev & Mendes, 2002) have shown that network connections have power-law distributions. That is, $p(k) \sim p^{-\gamma}$, where γ reflects some innate characteristics (Bonacich, 1972; Dorogovtsev & Mendes, 2002; Valente, Coronges, Lakon, & Costenbader, 2008). One of the most famous power-law-based examples is the Pareto principle, which states that 80% of events can be explained by 20% of the available causes. For word frequencies, Zipf's law

(Zipf, 1929) has been empirically observed by several researchers, who found that the word frequencies of corpora follow a power-law distribution (Kucera, 1985; Rousseau & Zhang, 1992). These studies show that it is not a new concept to explain data representativeness by network stability in the field of network theory. Therefore, we assumed that "network stability" could represent the theoretical saturation of qualitative data. That is, if a network stability occurs, we can say that the collected sample size is enough to have the representativeness of a certain group.

A network stability represents how well its global connectedness can overcome perturbations (Csermely, 2006). Perturbations are caused by intrinsic and extrinsic noise, and often disturb network structures. Therefore, we should ensure a proper level of local dissipation so that we can discriminate signals from noise (Csermely, 2006). Several network analysis studies described the representativeness of sub-samples of the population (Costenbader & Valente, 2003; Granovetter, 1976; Kossinets, 2006). Granovetter (1976) suggested a sampling strategy for ensuring representativeness in social networks, Costenbader & Valente (2003) observed the stability of centrality measures by dynamically adding and removing nodes and edges, and Kossinets (2006) considered the effect of missing data on network-level statistics. These studies analyzed the similarities of sample networks and the original network, using correlation analyses of the centrality measures (Costenbader & Valente, 2003; Valente et al., 2008).

By adopting this approach, we investigated the representativeness of data by observing the correlations between sub-network and original network. Due to its scale-free nature, Pearson's correlation coefficient (*r*) can be used to compare and measure the relationships between two different data sets. A network was represented with the centrality measures (Freeman, 1978) of node-level statistics, and shape and cohesion measures of network-level statistics. At first, these network measures fluctuate as the structure of the network changes with the addition of more nodes, as there are small numbers of participants or sample sizes. However, these measures do not significantly vary after the network becomes stable. Determining the smallest sample size that achieves network stability is a novel data collection strategy, which attempts to maximize effectiveness while minimizing cost and effort.

3.2. Method

This chapter aims to suggest a methodology of evaluating the reliability of user expression data. The suggested research method is adopted for case studies: two transcripts from qualitative research interviews and one set of customer review data on audio devices from Amazon.com (McAuley & Leskovec, 2013). This section introduces the used datasets and the research procedure used to identify network stability.

3.2.1. Datasets

For the case study research, we used datasets obtained from semi-structured interview transcripts regarding interactions with a robot and smart TV use. In addition, we collected customer reviews from Amazon.com to explore the possibility of adopting the suggested method for large datasets.

Dataset A contains interview transcripts describing users' emotional responses to the gestures of the humanoid robot MAHRU (Cha et al., 2011).

Three different movements were recorded for four gestural categories: expressive, symbolic, interactional, and referential. Twenty participants (male: 11, female: 9) in their twenties were asked 72 questions, such as “How did you feel after seeing this video? Give the reason for the emotion.” From the two-hour interview, the emotional responses of the participants to the robot’s movements with regard to its parts (e.g., hand, head, or shoulder) and the movement types (e.g., fast or slow) were collected. The results of the different gesture categories are separately analyzed in this paper. For convenience, each category is denoted by A1–A4, representing expressive, symbolic, interactional, and referential gestures, respectively.

Dataset B contains interview transcript data regarding four methods of interacting with a smart TV: pointing with a remote control, speech recognition, dialogue, and hand/arm gestures. Twelve participants (male: 6, female: 6) with an average age of 28.5 described their feelings when using a remote control for pointing, using speech recognition for unidirectional control, using dialogue to interact bi-directionally, and using hand and arm gestures. Before experiencing each interface, participants freely talked about their emotions toward the four types of interfaces. After completing the tasks, the participants were again asked about their feelings during interaction. The interview took about two hours, with 12 questions on the dialogue interface, 16 on the gestural interface, 12 on the remote control, and 12 on the speech recognition interface. During analysis, eight categories were separately analyzed: B1 and B2 represent the interview data before and after using the remote control, B3 and B4 represent the interview data before and after using the speech recognition interface, B5 and B6 represent the interview data before and after using the dialog interface, and B7 and B8 represent the interview data before and after using the gestural interface.

Dataset C is included because of the limited sizes of sets A and B. One thousand reviews on 20 speakers and 24 headphones/earphones on Amazon.com were randomly sampled. For these transcripts, the word boundary was defined as one review.

3.2.2. Research process

The research process suggested in this paper is composed of three stages, “generate semantic network,” “evaluate number of resamplings,” and “compare similarity between subnetworks and original networks,” as shown in Figure 3.1. The detailed description of the research process is introduced below.

3.2.2.1. Semantic network generation

We first transformed the text data into a co-occurrence matrix, which is a symmetric matrix with each cell containing the frequency of the words appearing in the units. In preprocessing, the word boundaries of Datasets A and B were defined as the answer of one participant. For Dataset C, terms appearing in one review were considered related. After defining word boundaries, processes such as stemming, stop-word removal, and synonym recognition were implemented. To reduce the time and effort required in this process, we utilized an automation tool (Lee, Rhie, Kim, & Lim, 2014) that uses two types of open source projects of the Lucene Korean Analyzer (2012) for Korean and Stanford’s CoreNLP (Toutanova, Klein, Manning, & Singer, 2003) for English. By utilizing this tool, we divided sentences into morphemes and eliminated stop words and terms that appeared less than five times. After generating co-occurrence matrices, we generated semantic networks using UCINET 6.0

(Borgatti et al., 2002) and Gephi (Bastian, Heymann, & Jacomy, 2009). Since one network represents the result of one participant, the result of *n* participants is obtained when *n* networks are merged (S_n), by adding nodes and edge weights of individual networks.

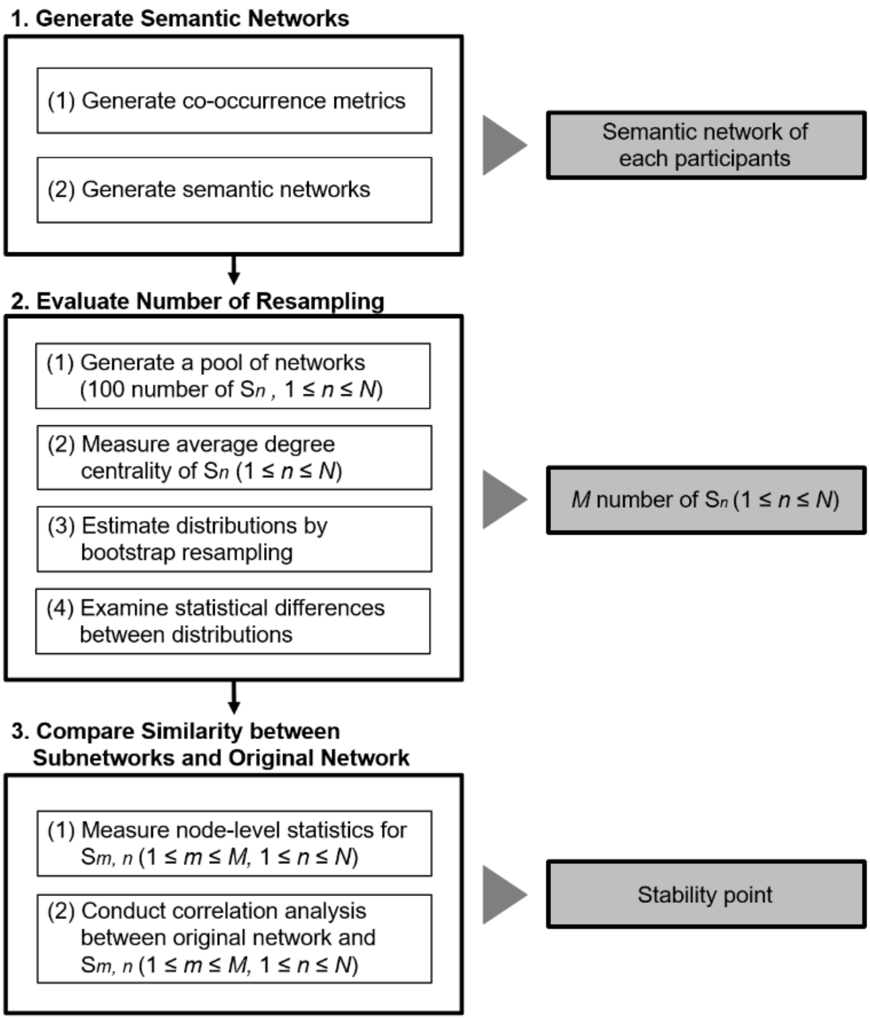


Figure 3.1. Procedure to identify stability points

3.2.2.2. Number of resamplings evaluation

To compare the networks of n participants ($1 \leq n \leq N$) with the original network, we had to predetermine the number of n-size networks to be observed. Therefore, we evaluated the requisite number (M) of subnetworks (S_n) to eliminate the effect of sample variability.

First, we formulated sufficient networks by sampling with replacement. We generated 100 sampled networks, which we assumed to be sufficient considering the study by Costenbader and Valente (2003). Then, we observed the network characteristics as we increased the number of resamplings. In this step, we used average degree centralities (average number of links per node) as the representative value of the formulated networks, since average degree centrality is highly correlated with the other node-level statistics (Valente et al., 2008). In addition, degree centrality directly reflects the word frequency, which has been used to determine the representativeness of corpora (Biber, 1993; Rayson & Garside, 2000) as well as the depth of the Internet review data (Huang & Yen, 2013; Mudambi & Schuff, 2010).

Based on the average degree centralities, we conducted bootstrap resampling, which is a statistical method that estimates sampling distribution using the Monte Carlo method (Hall & Martin, 1988). During bootstrap resampling, m representative values were randomly chosen (number of resamplings) 10,000 times (bootstrap replications); therefore, normal distributions of average degree centralities were estimated as increasing m. Then, the statistical differences between distributions were examined by Kullback–Leibler (KL) divergence (Kullback & Leibler, 1951), which calculates the amount of information loss between two distributions (equation 3.1).

$$I^{\mathrm{KL}}(F_0;F)=\frac{1}{2}\left[\frac{(\mu-\mu_0)^2}{\sigma^2}+\frac{\sigma_0^2}{\sigma^2}\right]-\log\left(\frac{\sigma_0^2}{\sigma^2}\right)-1 \tag{3.1}$$

If the distributions of m and m+1 showed information loss less than 0.05 bit, we set m as the minimum number of resamplings and denoted it as M. The process of identifying the number of resamplings is summarized below, and the logic is illustrated in Figure 3.2.

1) Sample 100 networks (s_n) with replacement for each sample size n = 1, 2, …, N, and calculate average degree centralities.

2) Among the set of average degree centralities, randomly sample m values 10,000 times (m = 1, 2, …, M).

3) Estimate the distribution of average degree centralities by a 5% confidence interval (CI) calculation. Since we sampled the network's degree centrality 10,000 times, this normality condition is satisfied.

4) By increasing m, the loss of information between samples $s_{n,m}$ and $s_{n,\mathrm{m}+1}$ is tested using the KL divergence.

5) Evaluate the number of sets (m) that ensures that each sample is representative, denoted as M.

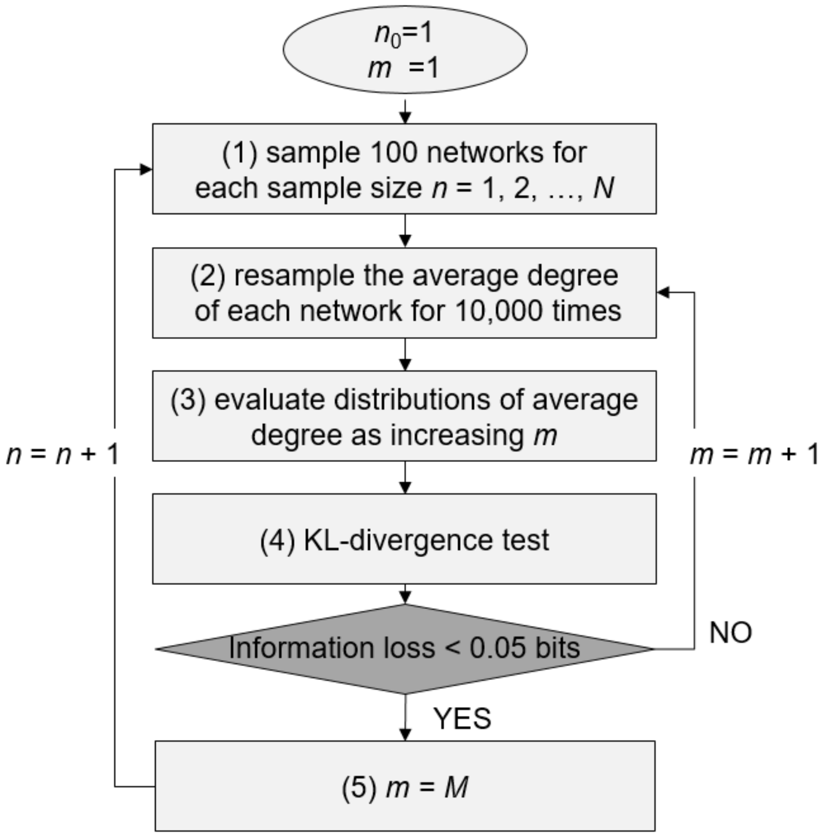


Figure 3.2. Method for determining the number of resamplings

Since it is time consuming to repetitively generate 100 networks, calculate average degree centralities, and estimate CI, we developed a tool using Gephi (Bastian et al., 2009) and R software. The pseudo code of this tool is demonstrated below.

MakeSample

read all text files (each text file is an interview transcript of individuals)

loop *1* : *100* **do**

Randomly sample *n* number of text files without resampling

write dataset as GML graph file #.gml

end loop

CalcCentrality

read all GML files (#.gml)

loop *1* : *number of GML files* **do**

Generate semantic network using Gephi toolkit

Calculate degree, betweenness, eigenvector centralities

write Calculated value to Centrality_#.txt file

Calculate average values of degree, betweenness, eigenvector centralities for each GML file

end loop

write average values of degree, betweenness, eigenvector centralities into the Output_Average.txt

MakeRScript

read average values of Output_Average.txt

write R command for bootstrap resampling on average degree centrality to Output_RScript_Degree.txt

Content of R command:

Sample *M* number of average degree centralities 10,000 times

Calculate confidence interval as increasing *M*

Write 95% confidence intervals of average degree centrality

3.2.2.3. Stability point evaluation

After obtaining a sufficient number of networks to represent sample size *n*, we examined the stability of networks by comparing the similarity between the node-level statistics of *n*-size networks and the original network. Because semantic network analysis combines mental models of users with quantitative calculations, the similarity between whole populations and sampled networks could be measured by conducting a correlation analysis on node-level centralities.

A correlation analysis was conducted on the centrality measures, which represent the importance and influence of a node in a network. We used the most frequently used centrality measures (degree centrality, betweenness centrality, and eigenvector centrality) to identify user values (Kim, Lim, Choi, & Yun, 2012).

We evaluated the Pearson correlation coefficients (*r*) to compare the size-*n* networks with the original. Figure 3.3 shows the correlation analysis procedure that compares the centrality measures of sample $s_{n,\mathrm{m}}$ ($n = 1, 2, \ldots, N$; $m = 1,2, \ldots, M$) and the original network. Then, the average of the correlation coefficients was calculated by $\mathrm{corr}(s_n) = M^{-1}\sum_{m=1}^{M}\mathrm{corr}(s_{n,m}, s_N)$. Finally, we observed whether stability occurred as *n* increased. Here, "stability points" correspond to the "saturation points" of theoretical sampling suggested by Glaser et al. (1968).

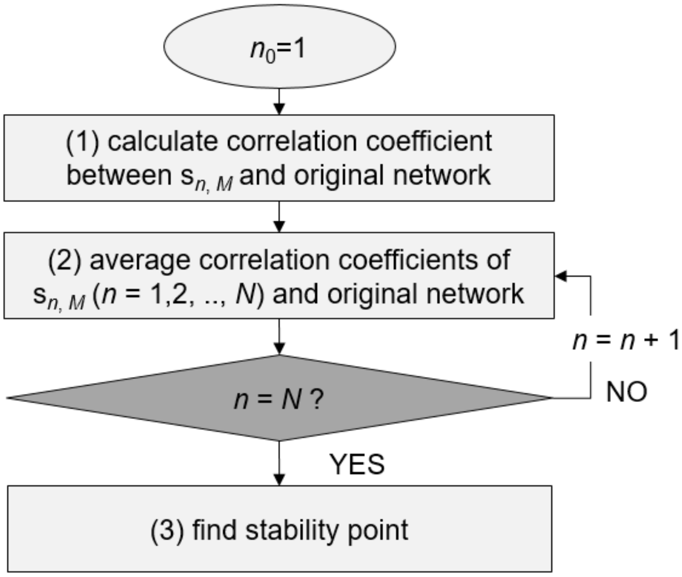


Figure 3.3. Method for evaluating the stability points

3.3. Results

After deriving co-occurrence metrics, we analyzed the network statistics to generate relational matrices between the sample and original networks. The interview data differed in features such as word count and data quality. In this study, the network stability was assessed in relation to its characteristics.

3.3.1. Descriptive statistics and network-level statistics

Table 3.1 illustrates the characteristics of the collected data. A further analysis of Dataset A was conducted for gestural categories A1–A4. On average, 658 nodes for each category, 7.336 nodes for each question, and 75.150 nodes per participant were collected. In Dataset B, 346 nodes were generated for each interface. On average, 7.815 nodes were generated per question and 45.396 nodes were collected for each participant. Dataset C yielded 1565 nodes, with an average of 18.378 generated per review. These three datasets were collected in various environments and on heterogeneous subjects, so a wide range of word counts was revealed. This was especially true for Dataset C, because the result of Dataset C showed much difference from Datasets A and B. As Dataset C was characterized by its written environment, which lacked interference from interviewers and specific questions while the reviews were composed, the number of nodes between participants showed a large variance.

Table 3.2 contains the network-level statistics for the semantic networks, which represent the network characteristics of the semantic networks. For example, Figure 3.4 illustrates the semantic networks for Categories B7 and B8 from Dataset B, with B7 exhibiting a simpler and more radial shape. Here, cohesion and connectivity are quite similar, but B7 has a higher clustering coefficient (2.036 compared with 1.884 for B8). The shape measures also differ in terms of the average distance (2.340 for B7 and 1.964 for B8) and diameter (4 for B7 and 2 for B8).

Table 3.1. Description of the interview data

		word count	average word count/questio	average word	std. word count between
Dataset A	A1	610	3.525	64.150	15.560
	A2	881	9.413	109.400	29.496
	A3	375	4.008	40.300	10.628
	A4	322	4.000	29.250	5.665
Dataset B	B1	171	8.944	19.583	10.672
	B2	389	7.613	52.083	11.453
	B3	233	7.569	30.250	10.154
	B4	343	7.594	45.167	9.124
	B5	478	8.944	62.667	30.338
	B6	381	7.813	48.916	11.704
	B7	409	7.042	52.250	12.607
	B8	364	7.000	52.250	12.607
Dataset C		1565	18.378	18.378	24.183

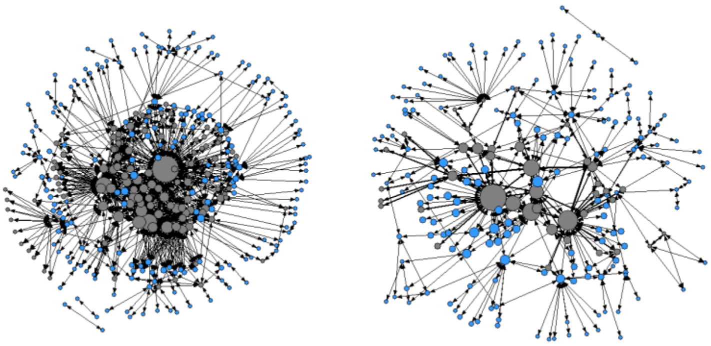


Figure 3.4. Networks from Dataset B

Note: hand and arm gestures before (category B7, left) and after (category B8, right). The node size is proportional to the eigenvector centrality; gray denotes tie strengths over 3, and blue indicates tie strengths over 1 and less than 3.

Table 3.2. Network-level characteristics for interview data

		Shape				Cohesion			
		Ave. distance	Diameter	Breadth	Compact-ness	Ave. degree	Density	Connec-tivity	Clustering coefficient
Dataset A	A1	3.074	6.000	0.700	0.300	7.730	0.013	0.852	1.184
	A2	2.152	4.000	0.515	0.485	27.522	0.031	1.000	7.494
	A3	3.322	8.000	0.684	0.316	7.859	0.021	0.937	0.992
	A4	2.158	4.000	0.513	0.487	31.218	0.041	1.000	3.781
Dataset B	B1	2.368	4.000	0.534	0.466	15.509	0.091	0.988	1.496
	B2	2.639	6.000	0.605	0.395	14.147	0.036	0.949	1.228
	B3	2.292	4.000	0.545	0.455	15.266	0.066	0.958	1.622
	B4	1.950	2.000	0.484	0.516	17.830	0.050	0.983	2.045
	B5	2.269	4.000	0.540	0.460	21.933	0.046	0.975	2.631
	B6	1.961	2.000	0.486	0.514	14.870	0.039	0.990	1.878
	B7	2.340	4.000	0.551	0.449	13.765	0.034	0.985	2.036
	B8	1.964	2.000	0.488	0.512	13.147	0.036	0.989	1.884
Dataset C		1.972	4.000	0.465	0.535	148.997	0.095	0.997	4.684

As shown in Figure 3.4 and Table 3.2, the network measures illustrate the characteristics of a dataset. The relationships between word frequency and network-level statistics were estimated using Kendall's τ. The results indicated that the total word frequency was positively correlated with the average degree ($\tau = 0.606$; $p = 0.004$) and density ($\tau = 0.416$; $p = 0.050$) at the 5% significance level, whereas the average word frequency per person was negatively correlated with density ($\tau = -0.597$; $p = 0.005$). This shows that an increase of the word frequency corresponds to an increase of the average degree and density, while density increases when there are the common terms that participants mention frequently rather than when the terms are sporadically mentioned.

3.3.2. Number of resampling

To evaluate the number of resamplings, which ensures the representativeness of *n* participants, we resampled the *m* average degree centralities with 10,000 replication times on each sample size and calculated their distributions. As illustrated in Figure 3.5, the CI at the 5% significance level diverges at a certain resampling level, indicating it would be efficient to determine the minimum required number of resamplings. To examine information loss between two distributions, we used the KL divergence test by increasing *m*. The information loss of *m* compared with m+1 resamplings was set to be less than 0.05 bit. By identifying the appropriate number of resamplings, we can generate representative networks for each sample size more efficiently and compare the results with the original network.

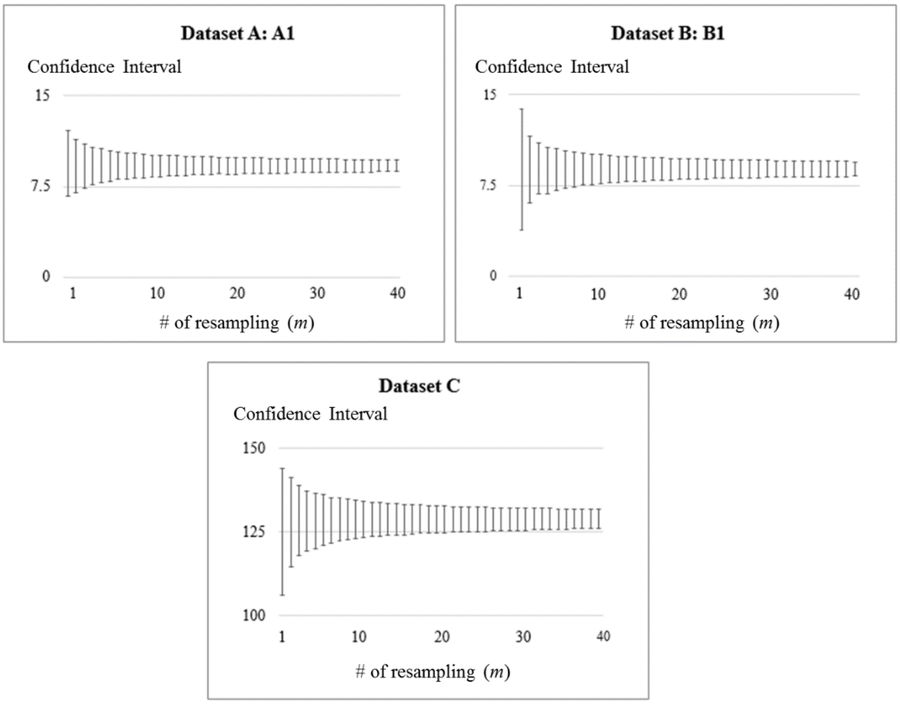
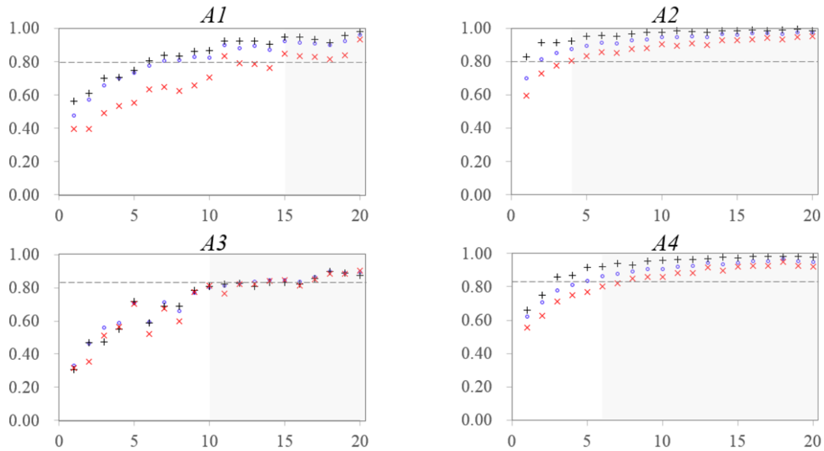


Figure 3.5. Confidence intervals versus number of resamplings

3.3.3. Network stability analysis

To determine stability points, we constructed and analyzed a relational matrix between each sample size and the original network. As illustrated in Figures 3.6–3.8, which represent the average correlation coefficients between the node-level centralities of subnetworks and the original network, the graphs converge to 1 as the sample size increases. Of the three centrality measures, betweenness has the highest correlation coefficient value, and eigenvector centrality has the lowest. This indicates that keywords bridging different groups (betweenness centrality) can be obtained with a smaller sample size compared to the popular or powerful keywords (betweenness centrality, eigenvector centrality).

We assumed that network stability is achieved when the correlation coefficient exceeds 0.800, which is a frequently used criterion for correlation coefficients. The result of Dataset A is illustrated in Figure 3.6, where the area above the dashed line indicates that the sample size reached stability. For category A1, the correlation coefficients of the degree, betweenness, and eigenvector centrality stabilized for sample sizes *n* of 7, 6, and 15, respectively. This implies that stability of A1 occurred when there were at least 15 participants. Similarly, the semantic networks of Datasets A2, A3, and A4 became stable at sample sizes of 4, 10, and 6, respectively. This means that we could use 50% or less of the participants to represent the results of 20 participants for the cases of A2, A3, and A4.



	A1	A2	A3	A4
stability point	15	4	10	6

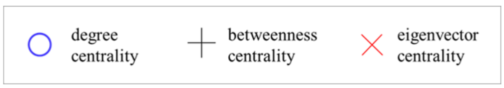
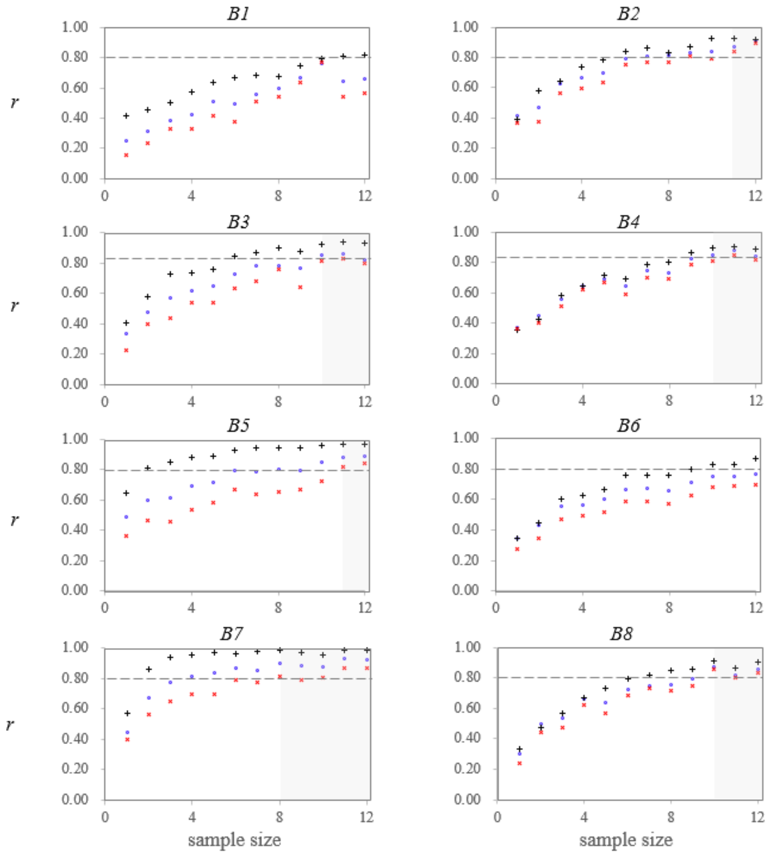


Figure 3.6. Stability points of Dataset A

Figure 3.7 shows the network stabilities for categories B1–B8 in Dataset B. B1 and B6 could not achieve stability, because the degree and eigenvector centralities showed low correlation coefficients. We assume the reason for B1 was that the participants mentioned small number of words, as shown in Table 3.1. For B6, referring low values of average distance and network density in Table 3.2, there were lack of pivotal words that connect the other important nodes. On the other hand, B7 became stable at a relatively small sample size compared to the other networks, as its correlation coefficients exceeded 0.800 for sample sizes of 8, whereas the other networks were stabilized at sample sizes of 10 and 11.

Dataset C was collected from unsolicited Internet reviews of audio devices on Amazon.com. As there was no interference or control of researchers when the transcripts were created, the number of words varied between the participants, as already shown in Table 3.1. The left panel of Figure 3.8 shows the overall network stabilities, revealing that the correlation coefficient significantly increased at around n =100. Therefore, we magnified a plot in the right panel to show that relatively few participants were required to appropriately represent 1000 people. The correlation coefficients r reached 0.700 when sample sizes were near 80, 50, and 120 for degree, betweenness, and eigenvector centralities, respectively. Furthermore, the correlation coefficients reached 0.800 at around 160 for degree centrality, 110 for betweenness centrality, and 220 for eigenvector centrality. From the result, it is reasonable to say that researchers should examine much larger sample size when using Internet review data, compared to the data collected by a structured interview technique. In this case, text-mining techniques such as sentiment analysis and opinion mining (Liu & Zhang, 2012; Zhuang, Jing, & Zhu, 2006) can be utilized to analyze abundant amount of textual data.



	B1	B2	B3	B4	B5	B6	B7	B8
stability point	-	9	10	10	11	-	8	10

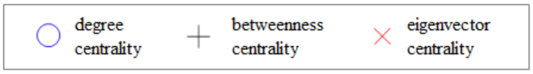
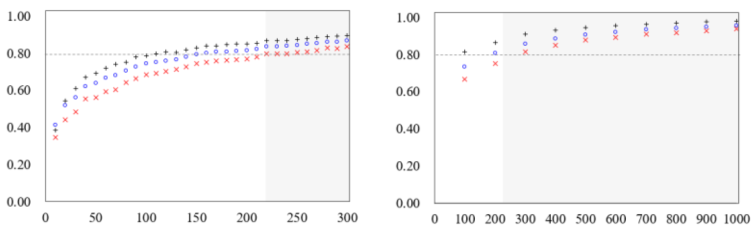


Figure 3.7. Stability points of Dataset B



	C
stability point	220

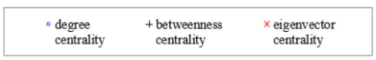


Figure 3.8. Stability point of Dataset C

3.3.4. Relationship between network characteristics and stability

To examine the relationship between network characteristics and stability, we determined whether there were mean differences between stability points for different network-level statistics by Wilcoxon signed-rank test (Wilcoxon, 1945) to Datasets A and B (Dataset C could not be concatenated due to the large sample size). From the analysis, we found that word frequency and network-level statistics had a significant effect on the stability point ($p<0.05$).

To identify whether there was a linear relationship, a Kendall correlation analysis (Kendall, 1938) between the network characteristics and the stability points was conducted. The results indicated that for betweenness centrality, smaller sample sizes were required for datasets with higher total word frequencies ($\tau = -0.457$), average word frequencies per person ($\tau = -0.476$), and clustering coefficients ($\tau = -0.457$). Eigenvector centrality also stabilized for smaller sample sizes in the presence of higher average degrees ($\tau = -0.580$), clustering coefficients ($\tau = -0.629$), and connectivities ($\tau = -0.619$) under the 5% significance level. In addition, different stability patterns of correlation coefficients were investigated, as some datasets showed radical increases at lower sample sizes, while others did not. To examine the relationship between the network characteristics and network stability, we observed the relationships between the quartiles of the correlation coefficients and the network-level statistics using a Kendall correlation analysis. The results indicated that variables of total word frequency, average clustering coefficient, and density were significantly correlated with the quartiles of stability points (Table 3.3). When observing the differences between the node-level statistics, eigenvector centrality was only affected by the density of the network, while degree and betweenness centralities were mainly affected by word frequency.

Table 3.3. Relationship between data description and network-stability quartiles (p<0.05)

	Quartile 1			Quartile 2			Quartile 3			Quartile 4		
	D.C	B.C	E.C	D.C	B.C	E.C	D.C	B.C	E.C	D.C	B.C	E.C
total word frequency	0.487	0.436			0.462		0.487	0.436		0.513	0.487	
ave.clustering coefficient								0.513			0.462	
density			-0.452	-0.452		-0.503						

Note: D.C, degree centrality; B.C, betweenness centrality; E.C, eigenvector centrality.

In the bottom quartile, the total word frequency was positively related to the degree and betweenness centralities, whereas the eigenvector centrality was negatively related to density (Table 3.3). In the second quartile, betweenness centrality had a positive relationship with total word frequency, and density was negatively related to the degree and eigenvector centrality. The third and the highest quartiles showed a similar relationship to the total word frequency's effect on degree and betweenness centralities and the average clustering coefficient's positive effect on betweenness centrality. Here, the effective variables of Quartiles 1 and 2 are important to identify the factors that induce stability to occur at a smaller sample size. Therefore, the results indicate that the total word frequency of data and density of a network are important to determine the stability of text data. This reveals that the absolute frequencies of terms have the predominant influence on a network's stability, but it is not a sufficient method to prove the stability.

Among network-level statistics, clustering coefficient has been alleged to indicate text quality (Antiqueira et al., 2007; Martin et al., 2013). Therefore, we analyzed the relationship between clustering coefficients and stability points. The result is illustrated in Figure 3.9 which demonstrates the relationships between the minimum percentage of the whole population and clustering coefficient of each dataset. In the figure, two datasets of A2 and C reaches stability points with small sample sizes for high cluster coefficients (Table 3.2). Considering higher clustering coefficients means lower text quality (Antiqueira et al., 2007; Martin et al., 2013; Pardo et al., 2006), the result suggests that achieving stability does not guarantee the quality of data, and vice versa.

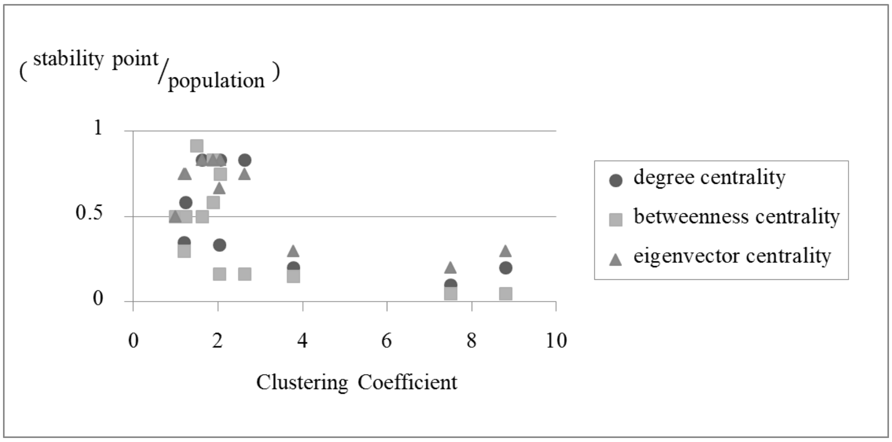


Figure 3.9. Relationship between appropriate sample size and clustering coefficient

3.4. Discussion

This study introduced the concept of network stability to assess the reliability of textual data in identifying UX issues. Based on a network-analysis approach, user expression data was presented with quantitative measures. During analysis, bootstrap resampling was conducted to derive representative subnetworks for each sample size, and correlation analysis was conducted to evaluate the similarity between *n*-size networks and the original network. To reduce the time and effort required for bootstrap resampling and network centrality evaluation, we developed a tool using the Gephi and R software packages. For case studies, the procedure was applied to two sets of in-depth interview transcripts and one set of online customer reviews. The main findings in this paper are described below.

First, different patterns of stability points were observed for different centrality measures. In general, eigenvector centrality had the stability point at the largest sample size, and betweenness centrality had it at the smallest. Such differences between centrality measures can be explained by their nature. Betweenness centrality indicates the shortest path between two nodes, rating zero for the terms that do not work as a bridge between two other words. Therefore, betweenness centrality counts fewer important words compared to the other node-level statistics. Degree centrality simply represents the number of connected nodes, in other words, the popularity of the words. Since there was no isolated node in our study, degree centrality in this study represented the word frequency. Meanwhile, eigenvector centrality considers the power of directly and indirectly connected nodes in a network; therefore, the measure is recalculated for every marginal node, which causes the perturbations of a network in the perspective of network theory.

Therefore, researchers can apply different statistics depending on their purpose; betweenness centrality can be used to represent important terms for information flow, while degree centrality should be used if the researcher wants to capture keyword categories, considering that word frequency has been used to measure the representativeness of a sample corpus in the field of linguistics (Biber, 1993; Rayson & Garside, 2000). Through eigenvector centrality, the concepts and relations of an original network can be identified when it reaches a stability point.

Second, we were able to observe the distributions of network stability as sample sizes increased. Figures 6 – 8 show a general pattern of network stability, which rapidly increases at a low sampling level and gradually increases from certain points. This is not a new finding, as semantic networks follow a general principle called the small-world structure, whereby a small number of nodes serve as hubs, while the node connectivities follow a power-law distribution (Steyvers & Tenenbaum, 2005). The power-law distribution of degree centrality has been reported in several studies (Bonacich, 1972; Dorogovtsev & Mendes, 2002; Valente et al., 2008). One of the power-law examples is the Pareto principle, which states that 80% of events can be explained by 20% of the available causes, which has also been observed for word frequencies (Kucera, 1985; Rousseau & Zhang, 1992; Zipf, 1929). For a certain dataset of case studies, 80% of the original network could be explained with less than 20% of the whole population: Categories A2 and A4 required only 2 and 4 participants among 20 participants, and Dataset C required only 200 of the 1000 participants. Although a larger portion of participants was required to reach stability points overall (50.9%, 42.1%, and 61.1% of all participants were required to ensure stability for degree, betweenness, and eigenvector centrality respectively), this study demonstrated that a smaller sample size could represent the majority of a dataset.

Third, we demonstrated a well-understood fact that a small number of participants can be adequate to obtain stability, as Crouch and McKenzie (2006) also argued that a small sample size is only sufficient if the researcher deeply explores the interviewees. Considering network-level measures imply text quality, we observed the effects of network-level statistics on stability points. The results indicated that the individual quartiles of the stability graphs were affected by different factors. In the first and second quartiles of correlation coefficients, centrality measures were positively affected by total word frequency (degree centrality, betweenness centrality) and negatively affected by density (degree centrality, eigenvector centrality). In the third and fourth quartiles, total word frequency (degree centrality, betweenness centrality) and average clustering coefficients (betweenness centrality) were positively correlated with stability points. This indicates that networks evolve more radically at lower sampling level, when there are higher word frequency and lower density. After forming the core network, the sampled networks become more akin to the original network when total word frequency and average clustering coefficients are higher. In summary, networks with higher word frequency, lower density, and higher clustering coefficients required smaller sample sizes to reach network stability. If relationships between network-level statistics and stability points are investigated for more various datasets, appropriate sample size will able to be evaluated without original network. Considering clustering coefficient is negatively related to the text quality (Antiqueira et al., 2007; Martin et al., 2013; Pardo et al., 2006), it should be noted that a low stability point does not mean that the data is high in quality. Therefore, during analysis, researchers need to examine the contents of text data even if the networks reach stability points.

It should be noted, however, that we used a limited number of statistical measures as they were the most frequently used. Other statistical measures such as PageRank centrality could be used to evaluate the network stability. Additionally, we used bootstrap resampling for a more efficient analysis, which inevitably leads to information loss. Despite these limitations, we have shown that the representativeness of qualitative data can be investigated at the semantic level. Additionally, the different tendencies of the network stability were identified by observing the relationship between quartiles of correlation coefficients and network-level statistics.

To sum, this chapter suggested a method to calculate the representativeness of qualitative interview data; therefore, researchers could avoid subjectivity with less time and effort in judging the stability of textual data. Moreover, by utilizing our method, UX researchers and practitioners would be able to collect the optimal sample size by gradually increasing sample sizes. We also presented three case studies composed of two interview datasets and one online review dataset, which proved that this method could be adopted for small as well as large sample sizes. In further research, we hope that the proposed procedures could be applied to automatic processing during the interview process, and help researchers quantitatively measure the reliability of their sample sizes.

CHAPTER 4

Identifying User Values using Qualitative and Quantitative Data for Camera Shutter Sounds

4.1. Overview

This section introduces a mixed method research approach (Teddlie & Tashakkori, 2009) that uses both qualitative text data and quantitative data in identifying UX. Effective user values (UV) were identified based on qualitative text data, and their relations to product attributes (PA) were explored by quantitative study. In this chapter, customer satisfaction on camera shutter sounds were explored; important UVs were selected by collecting and analyzing text data, and their relations to PA were identified by statistical analysis. Since it is difficult for general users to describe PA of camera shutter sounds, the concept of psychoacoustic variables was adopted to bridge these variables.

Whereas “intentional sounds” produced from a speaker or piezo elements have been created to improve certain impressions, “consequential sounds” emitted by the structure of a product depend on its mechanical properties (Langeveld, Van Egmond, Jansen, & Özcan, 2013). It is generated from physical factors transmitted to psychoacoustic factors that induce psychological reactions in customers which influence their attitudes and behaviors, such as preference for industrial products (Özcan & Schifferstein, 2014; Petiot, Kristensen, & Maier, 2013; Pietila & Lim, 2012; Västfjäll, Gulbol, Kleiner, & Gärling, 2002), food-related behaviors (Elder & Mohr, 2016; Spence & Shankar, 2010; Zampini & Spence, 2004), and brand image enhancement (Fastl, 2005; Flath & Klein, 2014; Lyon, 2003).

Therefore, companies have made efforts to develop appropriate sounds to improve affective reactions from customers. They have frequently used a basic affective circumplex of pleasantness-unpleasantness and activation-deactivation (Bradley & Lang, 2000; Russell, 1978; Västfjäll et al., 2002) or other affective variables such as powerfulness for cars (Bisping, 1997), gentleness for washers (Bowen & Carow, 2001), and strong feeling for buttons (Ishimitsua et al., 2008), depending on the research purpose. One of the most successful examples is Harley-Davison motorcycles, which established a distinctive brand identity using its powerful engine sound (Pierson & Bozmoski, 2003). Instead of PA, existing researchers measured perceptual characteristics of the targeted sound and conducted a correlation analysis with affective reactions of users. Here, the concept of psychoacoustics was adopted to represent perceptual characteristics.

Traditionally, PA of a consequential sound means the material property or structure of a product (Wang, 2015). For example, shutter speed is one of the PAs which determines duration of a shutter sound. However, sound engineers these days become able to shape, or even remove the existing sounds. One example is an electric vehicle which produces much less noise than traditional combustion engine sounds. In addition, mirrorless cameras also can eliminate the mirror and optical viewfinders which generate the typical shutter sounds from SLR (Single-lens reflex) and DSLR (Digital single-lens reflex) cameras when taking a picture. Since such consequential sounds work as indicators that give certain feedback (rather than being silent), product developers are trying to modify the sounds, rather than replace them. Therefore, it has become necessary to design consequential sounds rather than adjust the mechanical structure. In this paper, as time structure of shutter sounds can be modified, tempo-related variables, such as duration, were considered as PA instead of mechanical property of cameras.

The objective of the present chapter is to identify important UVs, and evaluate the effects of PA of camera shutter sounds. During research, we focused on adjusting time structure-related PA as the tonality was determined by the device's material properties. UVs were identified by users' descriptions on "satisfaction" of camera shutter sounds, and the relationships between PA and UV were analyzed by using psychoacoustic variables.

4.2. Measure

It is widely appreciated that physical property of a product does not imply what human perceives. The simplest indication of this phenomenon is Weber's law that people cannot discern between two objects when intensity is not changed for a certain amount (Fechner, 1966). Therefore, sensory perceptions such as vison, hearing, and touch are extracted with objective measures in the field of phsychophysics. Psychophysics identifies the relationships between the physical measurements of stimuli and perceptions of a human. It has been used to derive quantitative measures of perceptual phenomena that are considered subjective. Table 4.1 demonstrates examples of psychophysical variables for visual, auditory, and tactile stimuli (Fairchild, 2013; Okamoto, Nagano, & Yamada, 2013; Zwicker & Fastl, 2013).

Table 4.1. Psychophysical variables for different types of stimuli

Stimulus type	psychophysical variables
visual	illumination observer acuity eye movements eye fixation duration number of eye fixation complexity observer acuity
auditory	loudness sharpness roughness fluctuation strength
tactile	friction hardness warmness fine roughness macro roughness

Auditory perception is explained by psychoacoustic metrics (Blauert & Jekosch, 1997). PA of a sound is the spectral-temporal composition (varying the frequency spectrum and time structure of a sound), spectral features (the bandwidth and line spectra of a sound) (Fastl, 2006), and the time structure (the duration and arrangement of signals and pauses from the rise and decay of the peak). Meanwhile, psychoacoustic variables represent the sensorial reactions of a human (Zwicker & Fastl, 2013). The measures intuitively represent how people perceive the sounds by considering the limitation of human ear.

Many studies on sound quality have analyzed the effects of psychoacoustic variables on UV. For example, Västfjäll, Kleiner, and Gärling (2003) revealed that in the case of interior aircraft sound, valence was primarily related to loudness and activation was related to sharpness. Moreover, Zwicker and Fastl (1999) predicted sensory pleasantness with four parameters: loudness, sharpness, roughness, and tonality as follows:

$$\frac{P}{P_0} = e^{-\frac{0.7R}{R_0}} e^{-\frac{0.108S}{S_0}} (1.24 - e^{-\frac{2.43T}{T_0}}) e^{-(0.023N/N_0)^2} \quad (4.1)$$

where P indicates pleasantness, S indicates sharpness, R indicates roughness, T indicates tonality, and N indicates loudness.

4.2.1. Loudness (*N*)

Loudness is defined as the subjective impression of sound intensity. While sound pressure level (SPL) indicates the physical strength of sound pressure, loudness refers to the subjective perception of sound frequency. The perceived loudness according to the sound pressure level and frequency is shown in equal-loudness contours suggested by ISO 226:223 (2003).

Loudness is expressed in “phon” and “sone”, where both units represent the perception of loudness while compensating for the effect of frequency. Whereas phon uses a logarithmic scale, the sone scale is linear. The relationship between the two psychoacoustic measurements is shown in equation (4.2).

$$N(\text{phon}) = 40 + 10 \log2(N(\text{sone})) \quad (4.2)$$

Loudness is one of the most influencing factors in evaluating product sound quality. Fastl (2000) evaluated the effect of loudness on user preferences of electric razor sounds. For car engine sounds, higher loudness reduces comfort and elicits annoyance (Horvat, Domitrovi, & Jambrošić, 2012; Västfjäll et al., 2002), although it increases sporty impression (Coen et al., 2004).

4.2.2. Sharpness A (*S*(*A*)), Sharpness Z (*S*(*Z*))

Sharpness is closely related to timbre and represents a person's feelings toward the dullness or sharpness of a sound. The sharpness unit is expressed in "acums", where 1 acum represents the sound of 1kHz of frequency, 60dB of sound pressure, and 1 critical bandwidth. Sharpness is quantitatively calculated as,

$$S(Z)\ (\text{acum})=0.11 \cdot \int_0^{24\text{Bark}} N'g'(z)z\ dz \,/\, \int_0^{24\text{Bark}} N'\ dz \qquad (4.3)$$

where the denominator indicates the total loudness *N*, the upper integral is the first moment of specific loudness (*N*') over the critical band rate, and a weighting function *g*(*z*) depends on sound frequency and is boosted when it is over 16 Bark. The sum of the weighted partial moments is multiplied by a constant ($c = 0.11$) (Zwicker & Fastl, 2013).

Another model to calculate sharpness was provided by (Aures, 1985; Cabrera, Ferguson, & Schubert, 2007) in equation (4.4).

$$S(A)\ (\text{acum}) =0.585 \cdot \int_0^{24\text{Bark}} \mathrm{N'(z)\cdot g(z)\ dz \,/\, ln\ [(N+20)/20]} \qquad (4.4)$$

The two models have major differences when handling frequency in that the Aures model emphasizes higher frequencies over 4.6 KHz and attenuates low frequencies, while the Zwicker model counts the medium band energies and instantaneous frequencies of the highest band.

Sharpness generally indicates a negative effect of sound quality (Nykänen & Sirkka, 2009; Wagner & Kallus, 2013; Zwicker & Fastl, 1999). In the case of impulsive sounds, Hoechstetter, Sautter, Gabbert, and Verhey (2016) identified that the duration of sharpness is more strongly correlated to sound quality than loudness. In the meantime, sharpness also helps to improve certain target feelings; Västfjäll et al. (2003) and Coen et al. (2004) revealed that sharpness gives positive effects on activation and sportiness, respectively.

4.2.3. Roughness (*R*)

Roughness represents a temporal variation of the sounds caused by amplitude fluctuation (Daniel & Weber, 1997). The unit is "asper", and 1 asper is defined as a pure tone sound with an amplitude of 60dB at 1kHz, calculated as

$$R \text{ (asper)} \approx \Delta \mathrm{L} \cdot f_{\mathrm{mod}} \qquad (4.5)$$

where ΔL is the modulation depth of temporal-masking pattern and f_{mod} is the modulation frequency (Zwicker & Fastl, 2013).

Roughness is described as buzzing and harsh sound, presenting unpleasant and annoying feelings (Coen et al., 2004; Horvat et al., 2012). Meanwhile, roughness gives positive effects on activation, which is positively related to user preference on interior aircraft sounds (Västfjäll et al., 2002).

Using these psychoacoustic variables of loudness, sharpness, and roughness, existing researchers have investigated the relationship between psychoacoustic variables and UV. Table 4.2 shows the effects of psychoacoustic variables on UV vary depend on the targeted product and context of use.

Table 4.2. Literature review on psychoacoustic analysis

Literatures	Stimuli	Effects of psychoacoustic variables	UV
Dedene et al. (1998)	car exhaust sound	Pleasantness and sportiveness are negatively related to loudness and sharpness.	pleasant, sporty
Västfjäll et al. (2002)	exterior, interior vehicle sound	Loudness negatively affects valence; roughness positively affects activation.	valence, activation, preference
Västfjäll et al. (2003)	interior aircraft sound	Activation is affected by sharpness, and valence is affected by loudness and roughness.	valence, activation
Coen et al. (2004)	car engine sounds	Loudness, sharpness, and roughness negatively affect comfort, and positively affect sportiness.	comfort, sportiness
Zampini and Spence (2004)	potato chip	Crispness and staleness are affected by loudness and frequency composition.	crispness, staleness
Lee (2008)	booming, rumbling sound	Loudness filtered by a low pass filter and sharpness have negative and positive relationship.	subjective rates (bad ~ excellent)
Nor et al. (2008)	interior vehicle sound	Loudness and fluctuation strength are negatively related to VACF, sharpness and roughness have positive relationship.	vehicle acoustical comfort index (VACF)
Nykänen and Sirkka (2009)	automobile power window	Low loudness, sharpness, and motor speed fluctuations lead high product quality.	sound quality, annoying, under-powered
Horvat et al. (2012)	vacuum cleaner	Annoyance index correlates loudness and sharpness.	annoyance
Wagner and Kallus (2013)	turn indicator sounds	Loudness, sharpness, impulsiveness is negatively related to sound quality.	sound quality
Hoechstetter et al. (2016)	vehicle door locking and closing sounds	Sound quality decreases as duration of sharpness increases.	sound quality

4.3. Research process

A sound is featured by its spectral and time structure, influencing perceptual characteristics of loudness, sharpness, and roughness. After perceiving a sound, people can describe what the sound is like, and which impressions they received. Therefore, we collected user's opinions (qualitative text data, satisfaction scores) on existing shutter sounds to identify UV and PA of camera shutter sounds. Then, their relationships were identified by statistical analysis. The research process of this chapter is organized by five phases (Figure 4.1): (1) Elicit PA of shutter sounds, (2) Collect qualitative text data and satisfaction score, (3) Elicit important UVs, (4) Evaluate effective PAs, (5) Modify camera shutter sounds, (6) Conduct jury test on modified shutter sounds. The research process is described with more detail in the rest of this section.

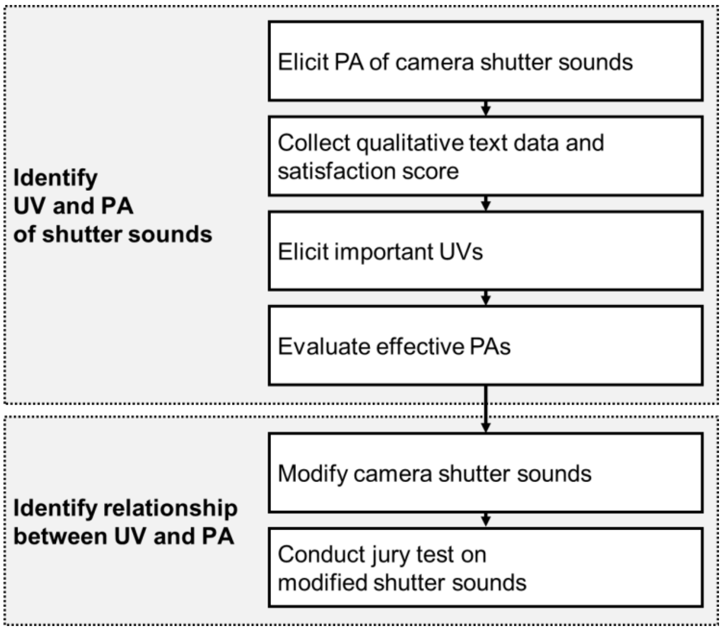


Figure 4.1. Procedure of research method

4.3.1. Eliciting PA of camera shutter sounds

The focus of this study is on how the time structure of sounds in the frequency spectrum is affected by the materials and shapes of the camera components. The features of the time structure of the camera shutter sound were investigated with two sound engineers of the SoundSketch Company in Korea.

Since typical shutter sounds are consequentially emitted from the upward and downward movements of the shutter and mirror, they have certain features contained in temporal envelopes. As illustrated in Figure 4.2, the envelopes are composed of “peaks” and “intervals” between them.

Based on ten camera shutter sounds, four UX researchers and two sound engineers specified PA that describe the sound. The qualitative variables describe the shape of the envelopes (decrease, increase, and uniform) based on the amplitude values of the first and last peaks, while the number of peaks, total duration, and peak-to-peak duration comprised the quantitative variables. Peak-to-peak duration was defined as the duration between the two largest peaks of the shutter sounds.

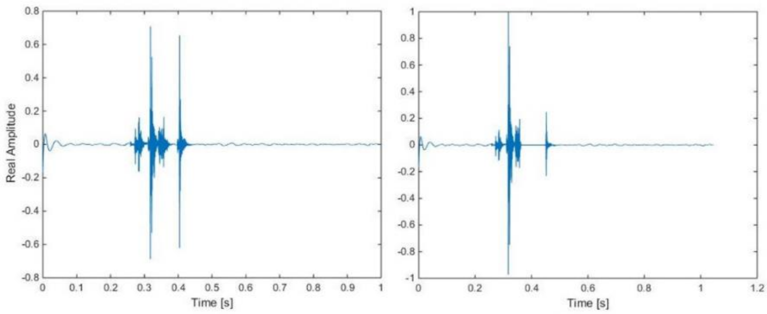


Figure 4.2. Time-varying amplitudes of shutter sounds

Note: Left side of the figure represents the stable envelope, and right side of the figure represents decreasing envelope

4.3.2. Conducting jury test on existing camera shutter sounds

To examine the impact of PA on satisfaction, a jury test was conducted to survey satisfaction scores on ten existing cameras (four DSLR cameras, five mirrorless cameras, and one compact camera). 50 participants from different demographic backgrounds (age - 20's: 27; 30's: 17; 40's: 6, gender - male: 25; female: 25) and camera experience (under 5 years: 22; over 5 years: 28) took part in the test. The cameras were set on tripods in a line, and the participants were asked to stand behind the cameras at a distance of 50cm, considering that people usually click shutters behind the cameras. Before experiment, participants wore an eye patch to prevent visual stimulation. After hearing camera shutter sounds in a random sequence, participants rated satisfaction scores in the basis of 100 point, and described reasons for the scores.

Figure 4.3. Experimental environments of the first jury test

4.3.3. Eliciting important UVs

Before generating a semantic network, preprocessing step was performed. This step includes defining word boundaries, stop words, and synonyms. We used an automation tool developed by Lee et al. (2014) that helps extracting co-occurrence matrix from natural language. First, the window of co-occurrence, in other words, word boundaries, was defined: the terms that came out from the same participants were considered to be associated. Second, the word stemming step which decomposes clauses into morphemes was proceeded. Lucene Korean analyzer (http://cafe.naver.com/korlucene) which has been mostly used in analyzing Korean, was utilized. Third, stop words such as *a*, *the*, and *to*, and other designated words were eliminated. In addition, synonyms, like *year* and *yrs*, are unified into the same letters. In this process, five UX researchers shared opinions to prevent involvement of subjective judgement.

Co-occurrences of two terms were encoded into an edge between the two nodes, while each node represents a term. Using the node and edge information, a relational matrix was obtained to calculate node-level statistics including degree, closeness, betweenness, and eigenvector centralities using UCINET 6 (Borgatti et al., 2002). The normalized centrality measures of the collected terms are listed in Appendix A. At last, keywords in a network were classified into three categories: UV, PA, and contextual variable (CV). CV implies the environmental factors of tasks, in this case, the physical parts of a camera. In this study, UV were selected from top 10 keywords of degree, betweenness closeness, and eigenvector centralities.

4.3.4. Evaluating effective PAs

The satisfaction scores of participants were transformed into $0 - 1$ range by min-max normalization. Demographic factors such as gender (p = 0.521), age (p = 0.658) and camera experience (p = 0.743) had no significant effect on satisfaction scores. Camera type also showed no significant influence except for the compact camera which received a significantly lower score ($p < 0.05$).

Users' satisfaction scores on PAs are presented in Figure 4.4. For total duration, satisfaction showed no linear relationship, and received their highest scores when the duration was less than 1,500 ms. For envelope shape, all three shapes had no significant differences ($p = 0.704$). For peak-to-peak duration, a duration of less than 106.7 ms received significantly higher scores ($p < 0.05$).

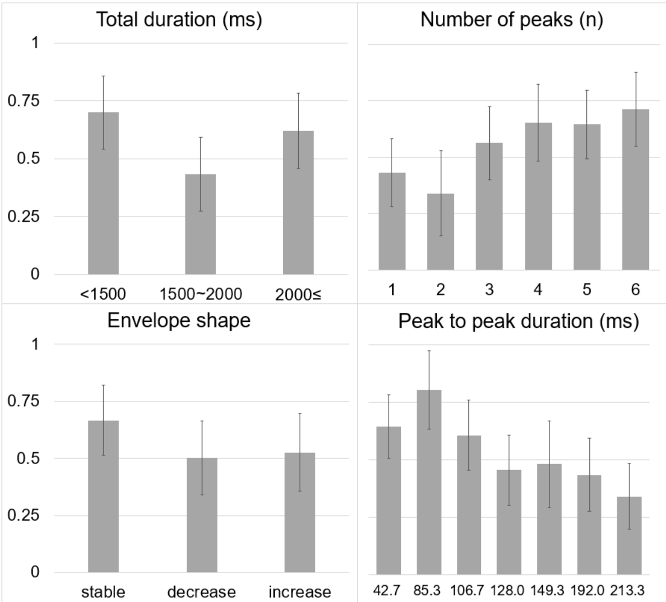


Figure 4.4. Effects of PA on satisfaction scores

4.3.5. Modifying camera shutter sound

For more thorough observations on the impact of the PA, we modified a shutter sound to control the effect of tonality. Before modifying the sound, we determined independent and control variables based on the first jury test. First, four PA, total duration was controlled to last 1,000 ms as sounds shorter than 1,500 ms received the highest score. Second, sounds were formulated to have more than three peaks as satisfaction received significantly higher scores when there were more than three peaks. Third, all three envelope shapes (stable, increase, and decrease) were applied. Since no significant effect was found in former experiment, we tried to assure under a controlled environment. Fourth, peak-to-peak duration was set at 80 ms, which received the highest satisfaction score, and 130 ms, which received relatively lower scores.

Among the composites of PA, we conducted a pilot test to reduce the number of sound samples. Ten participants heard a pair of sounds and were asked if they could identify differences between them. If they could, the more preferred sound was selected. During the process, sounds with five or more peaks were ruled out, as people felt they did not sound like a camera shutter sound. Therefore, only three to four peaks were chosen. Among the variations of peak-to-peak durations (130 ms, 80 ms), number of peaks (3 peaks, 4 peaks), and envelope shapes (stable, decrease, increase), four samples with the lowest preference were also excluded. As a result, the eight sounds listed in Table 4.3 were selected for further analysis.

Table 4.3. PA of modified sound samples

Sample	Time Structure			FFT Spectrum		
	peak-to-peak duration	number of peaks	envelope shape	spectral centroid	kurtosis	skewness
S1	130	4	stable	204.99	134.83	10.53
S2	130	4	decrease	200.82	138.64	10.87
S3	130	3	increase	202.70	130.64	10.55
S4	130	4	increase	204.99	130.05	10.51
S5	80	4	stable	187.74	139.03	10.93
S6	80	3	decrease	196.03	140.10	10.95
S7	80	4	increase	187.74	139.03	10.93
S8	80	3	increase	187.74	139.03	10.93

4.3.6. Conducting jury test on modified shutter sounds

To identify subjective ratings on sound samples, we conducted the second experiment on 30 participants. The demographic characteristics of age (20's: 16; 30's: 12; 40's: 2), gender (male: 16; female: 14), and camera experience (under 5 years: 13; over 5 years: 17) were approximately counterbalanced. The questionnaire was composed of two sections: eight items of UV, which were elicited in the first jury test, were evaluated using 7-point Likert scales, and satisfaction scores were assessed on the basis of a 100-point scale.

Before conducting the experiment, participants were introduced to the questionnaires and were then required to put on AKG K601 headphones to hear all of the sound samples. During the experiment, participants were allowed to hear each sound as much as they wanted before filling in the questionnaire items. The sequence of hearing order was randomly formulated for each participant.

Figure 4.5. Experimental environments of the second jury test

4.4. Results

4.4.1. User values (UV)

From semantic networks illustrated in Figure 4.6, centralities of the keywords were calculated and listed in a descending order. Degree centrality, Freeman betweenness centrality, closeness centrality, and eigenvector centrality were used to screen important keywords. The list of all keywords and centralities are shown in Appendix A.

From Figure 4.6, we can notice that the core part of the network is usually composed of UV, showing small number of PA and CV. The phenomenon is also observed in Table 4.4, as it shows only one PA (*tonality*) in top 10 – node centrality. Although abstract level of description on camera shutter sounds were translated into keywords such as *sound envelope*, *peak*, and *duration*, those keywords had lower centrality comparing to the other datasets.

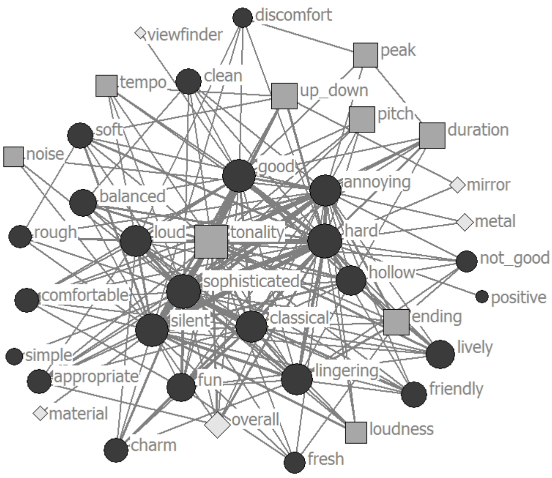


Figure 4.6. Semantic networks of collected text data

Note: The sizes of the nodes indicate their eigenvector centralities in the network. The dark nodes with a round shape are UV, the gray nodes with a square shape are PA, and the bright nodes with a diamond shape are CV.

Table 4.4. Top 10 centralities for camera shutter sound

Degree		Closeness		Eigenvector		Betweenness	
hard^{U}	0.851	hard^{U}	0.859	hard^{U}	0.307	hard^{U}	0.117
sophisticated^{U}	0.776	sophisticated^{U}	0.807	sophisticated^{U}	0.300	sophisticated^{U}	0.058
tonality^{P}	0.746	tonality^{P}	0.788	silent^{U}	0.296	good^{U}	0.057
good^{U}	0.746	good^{U}	0.788	tonality^{P}	0.293	tonality^{P}	0.056
silent^{U}	0.746	silent^{U}	0.788	good^{U}	0.292	loud^{U}	0.048
loud^{U}	0.701	loud^{U}	0.761	loud^{U}	0.281	annoying^{U}	0.045
annoying^{U}	0.687	annoying^{U}	0.753	annoying^{U}	0.281	silent^{U}	0.043
classical^{U}	0.672	classical^{U}	0.744	classical^{U}	0.277	classical^{U}	0.041
lingering^{U}	0.642	lingering^{U}	0.728	lingering^{U}	0.275	soft^{U}	0.036
hollow^{U}	0.582	hollow^{U}	0.698	hollow^{U}	0.257	fun^{U}	0.031

Note: Superscripts indicate keyword categories: U indicates UV, P indicates PA, and C indicates CV.

As shown in Table 4.4, there was no CV and only one PA in top 10 keywords, which might have come from the characteristics of an object, whose physical feature is difficult to be described without an external knowledge. Due to this characteristics, participants used an abstract term, *good*, more frequently; rather than saying “the *tonality* of a sound gave *classic* impression,” they said “the *tonality* was *good*”. Tidying synonyms and antonyms from Table 4.4, UV of shutter sounds were defined with eight adjectives: “hard,” “sophisticated,” “silent,” “annoying,” “classical,” “lingering,” “hollow,” and “fun”.

The result of network analysis was not much different from that of frequency analysis, considering 10 most frequently mentioned keywords were “hard (190),” “sophisticated (166),” “silent (128),” “tonality (127)”, “annoying (120),” “good (95),” “classical (90),” “lingering (52),” “hollow (52),” and “fun (52)”, while the numbers in the bracket implies frequency of each word.

4.4.2. User group identification

Before analysis, satisfaction score was transformed into 0-1 range by min-max normalization, and illustrated in Figure 4.7. The result demonstrates the highest score for S2 and lowest score for S4. As a result of t-test, gender (p = 0.477), age (p = 0.128) and camera experience (p = 0.767) had no significant effect.

Since standard deviations of satisfaction scores between participants were high with the highest value of 0.354, we conducted a cluster analysis to identify user groups. Statistical software package SPSS 18.0 was utilized to apply Ward’s minimum variance hierarchical clustering method, which is one of the most frequently used techniques in classifying subjects.

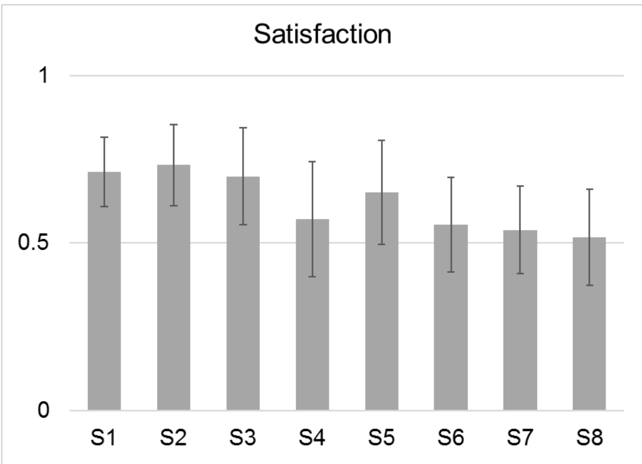


Figure 4.7. Normalized satisfaction scores of modified shutter sounds

As a result, four individuals were excluded as outliers, and two groups, denoted as G1 and G2 composed of 15 and 11 participants, were evaluated. The groups showed no significant difference between gender (p = 0.487), but revealed significant differences in age ($p < 0.001$) and camera experience ($p < 0.001$). The respective mean values of age and camera experience were 32.0 and 8.4 years for G1, and 28.5 and 5.2 for G2. As illustrated in Figure 4.8, the two groups show distinct differences in satisfaction scores. The results of an independent-sample t-test revealed that the two user groups had significant differences on S4 and S8 ($p < 0.001$).

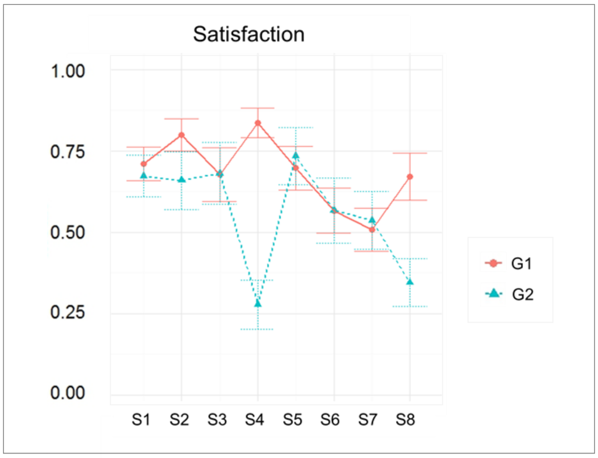


Figure 4.8. Satisfaction scores for G1 and G2

4.4.3. Psychoacoustic analysis of sound samples

The psychoacoustic measures of loudness (Moore, Glasberg, & Baer, 1997), sharpness (A), sharpness (Z) (Zwicker & Fastl, 2013), and roughness (Daniel & Weber, 1997) were calculated using Psysound3 software (Cabrera et al., 2007). Since psychoacoustic measures are time-varying, we used the average, minimum, 25%, 50%, 75%, and maximum as representative values during analysis. Their values are presented as a box plot in Figure 4.9.

Box plot illustrates thick tails of loudness, sharpness A, and roughness, as the values of Q1 and Q2 are much closer than those of Q2 and Q3. The loudness of S1 ~ S4, which have 130 ms of peak-to-peak duration, are higher than S5 ~ S8, which have 80 ms of peak-to-peak duration.

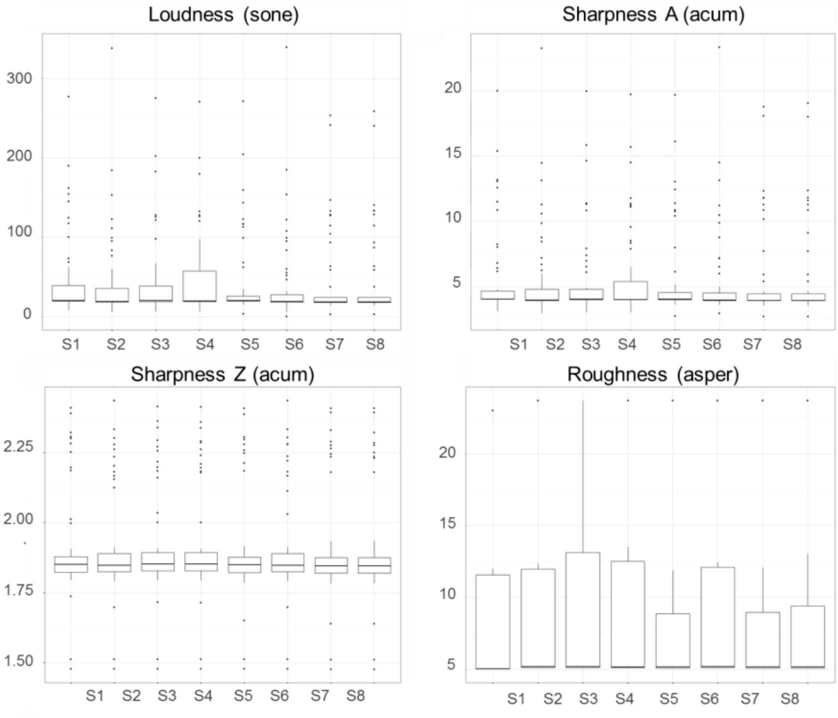


Figure 4.9. Psychoacoustic measures of modified shutter sounds

4.4.4. Regression model of user satisfaction

Since the objective variable, which becomes satisfaction level in this case, is difficult to be explained by psychoacoustic metrics (Västfjäll et al., 2003), we used eight UVs to relate the perceptual properties of shutter sounds, which were “hard” “fun”, “hollow”, “classical”, “annoying”, “lingering”, “sophisticated”, “silent”. The result of linear regression analysis is shown in the equations (4.6) and (4.7).

$$\text{Satisfaction (G1)} = 0.635 - 0.262 \times \text{annoying} + 0.250 \times \text{lingering} - 0.195 \times \text{hard} + 0.183 \times \text{sophisticated} \tag{4.6}$$

$$\text{Satisfaction (G2)} = -0.025 + 0.275 \times \text{silent} + 0.264 \times \text{lingering} + 0.234 \times \text{fun} + 0.298 \times \text{classical} - 0.206 \times \text{hollow} + 0.199 \times \text{sophisticated} \tag{4.7}$$

From the results, satisfaction models showed adjusted R^2 values of 0.582 and 0.488 (G1: equation (4.6), G2: equation (4.7)). As shown in the equations, user groups were influenced by different UVs; G1 showed a negative reaction toward “annoying” and “hard”, and positive toward “lingering” and “sophisticated”, while G2 showed a positive reaction toward “silent”, “lingering”, “fun”, “classical”, and “sophisticated”, and negative reaction from “hollow”. Both groups were commonly influenced by “lingering” and “sophisticated” in a positive way, but showed differences in that G1 was negatively influenced by “annoying” and “hard”, while G2 was negatively affected by “hollow”.

4.4.5. Effect of psychoacoustic variables on UV

Effective UVs on satisfaction were analyzed by psychoacoustic metrics. Table 4.5 shows the result of a correlation analysis for G1 along with the effect size of the regression model. Overall, weak but significant relationships were observed. These low level of correlation coefficients may be due to the small variations between sound samples.

However, we observed obvious relationship between psychoacoustic measures and UV. The G1 participants felt that sounds were less "hard" with higher N_{ave}, N_{min}, N_{Q3}, $S(\mathrm{A})_{ave}$, $S(A)_{Q3}$, $S(Z)_{ave}$, $S(Z)_{Q1}$, and $S(Z)_{Q2}$, which improved satisfaction. Meanwhile, they felt less "annoying" with higher N_{min}, N_{Q3}, N_{Q4}, $S(A)_{min}$, $S(A)_{Q3}$, $S(A)_{Q4}$, $S(Z)_{ave}$, $S(Z)_{Q1}$, $S(Z)_{Q3}$, R_{ave}, and R_{Q3}, which also improved satisfaction. However, the feelings of "sophisticated" and "lingering" showed no significant relationship with any of the psychoacoustic variables.

Table 4.5. Relationship between psychoacoustic measures and UV of G1

		hard	sophisticated	annoying	lingering
effect size (beta)		-0.246	0.231	-0.302	0.322
Loudness	N_{ave}	-0.225*	-	-	-
	N_{min}	-0.193*	-	-0.180*	-
	N_{Q3}	-0.291**	-	-0.231*	-
	N_{Q4}	-	-	-0.185*	-
Sharpness A	$S(A)_{ave}$	-0.238**	-	-	-
	$S(A)_{min}$	-	-	-0.241**	-
	$S(A)_{Q3}$	-0.296**	-	-0.242**	-
	$S(A)_{Q4}$	-	-	-0.185*	-
Sharpness Z	$S(Z)_{ave}$	-0.254**	-	-0.219*	-
	$S(Z)_{Q1}$	-0.204*	-	-0.216*	-
	$S(Z)_{Q2}$	-0.233*	-	-	-
	$S(Z)_{Q3}$	-	-	-0.221*	-
Roughness	R_{ave}	-	-	-0.234*	-
	R_{Q3}	-	-	-0.242**	-

Note: min., avg., Q1~Q4 represents minimum, average, and quartile values of the time-varying shutter sounds. * represents $p < 0.05$, ** represents $p < 0.01$.

Table 4.6 shows correlation coefficients between psychoacoustic measures and UV for G2. Participants in G2 were influenced by different UVs from those of G1. There were positive effects of “sophisticated”, “silent”, “sophisticated”, “lingering”, “fun”, and “classical” on satisfaction. In addition, there was a negative effect of “hollow”. The result of correlation analysis revealed that the participants felt sounds more “sophisticated” when the N_{Q3} and $S(A)_{Q3}$ decreased, and more “silent” when N_{ave}, N_{Q3}, $S(A)_{ave}$, $S(A)_{Q3}$, and $S(Z)_{ave}$ decreased. “Fun” was positively affected by N_{min}, and “classical” was positively affected by N_{Q1}, N_{Q2}, $S(A)_{min}$, $S(A)_{Q1}$, $S(A)_{Q2}$, and $S(Z)_{min}$. “Hollow” was not significantly influenced by the psychoacoustic measures.

Table 4.6. Relationship between psychoacoustic measures and UV of G2

		sophisticated	silent	lingering	fun	classical	hollow
effect size (beta)		0.241	0.294	0.308	0.248	0.300	-0.226
Loudness	N_{ave}	-	-0.314**	-	-	-	-
	N_{min}	-	-	-	0.230*	-	-
	N_{Q1}	-	-	-	-	0.240*	-
	N_{Q2}	-	-	-	-	0.257*	-
	N_{Q3}	-0.220*	-0.274*	-	-	-	-
Sharpness A	$S(A)_{ave}$	-	-0.318**	-	-	-	-
	$S(A)_{min}$	-	-	-	-	0.214*	-
	$S(A)_{Q1}$	-	-	-	-	0.217*	-
	$S(A)_{Q2}$	-	-	-	-	0.256*	-
	$S(A)_{Q3}$	-0.221*	-0.283**	-	-	-	-
Sharpness Z	$S(Z)_{ave}$	-	-0.220*	-	-	-	-
	$S(Z)_{min}$	-	-	-	-	0.270*	-

Note: min., avg., Q1~Q4 represents minimum, average, and quartile values of the time-varying shutter sounds. * represents $p < 0.05$, ** represents $p < 0.01$.

4.4.6. Effect of PA on psychoacoustic variable

In order to investigate the effect of the PA, we calculated the average values of the psychoacoustic metrics for each PA, as shown in Table 4.7. The results show that a longer peak-to-peak duration tended to have higher loudness and sharpness A, except for the minimum and maximum values (N_{min}, N_{Q4}, and $S(A)_{\mathrm{Q4}}$). For envelope shape, decreasing shapes showed lower loudness and sharpness A (N_{ave}, N_{Q1}, N_{Q2}, $S(A)_{\mathrm{ave}}$, $S(A)_{\mathrm{Q1}}$, $S(A)_{\mathrm{Q2}}$), and higher roughness. In addition, 4 peaks showed higher loudness (except for N_{Q4}) and lower roughness than 3 peaks. The relationships between PA, psychoacoustic variable, and UV are further discussed in the discussion and conclusions section.

4.5. Discussion

The aim of this study was to identify effective UVs and PAs that influence satisfaction of users, and their relationships in case of a camera shutter sound. Throughout this study, we were able to evaluate important UVs by applying semantic network analysis on qualitative text data. The terms were analyzed by non-hierarchical network structure, and important UVs were elicited by using degree, betweenness, closeness, and eigenvector centralities. The result of network analysis was not much different from frequency analysis, as our data had only 67 words, and had a cohesive network structure. However, network analysis will help evaluating important keywords that frequency analysis can’t find, if applied on larger datasets as centrality measure considers more information than words frequency.

Table 4.7. Relationship between PA and psychoacoustic variable

		peak-to-peak duration (ms)		envelope shape			number of peaks	
		80	130	stable	dec	inc	3peak	4peak
Loudness	N_{min}	2.32	1.64	1.64	2.84	4.65	3.18	3.61
	N_{ave}	43.70	45.26	45.26	41.86	45.01	43.34	44.85
	N_{Q1}	18.45	19.51	19.51	18.25	18.59	18.50	18.88
	N_{Q2}	18.92	20.07	20.07	18.85	19.12	19.03	19.44
	N_{Q3}	25.45	29.34	29.34	31.63	36.15	30.04	35.29
	N_{Q4}	280.38	273.95	273.95	338.56	264.31	290.89	281.92
Sharpness A	$S(A)_{min}$	2.78	2.93	2.93	2.93	2.86	2.88	2.90
	$S(A)_{ave}$	5.54	5.66	5.66	5.45	5.63	5.53	5.63
	$S(A)_{Q1}$	3.93	4.00	4.00	3.91	3.94	3.93	3.96
	$S(A)_{Q2}$	3.96	4.03	4.03	3.96	3.97	3.97	3.99
	$S(A)_{Q3}$	4.48	4.60	4.60	4.66	4.77	4.59	4.77
	$S(A)_{Q4}$	20.20	19.85	19.85	23.29	19.36	20.77	20.28
Sharpness Z	$S(Z)_{min}$	1.48	1.48	1.48	1.48	1.48	1.48	1.48
	$S(Z)_{ave}$	1.91	1.91	1.91	1.91	1.91	1.91	1.91
	$S(Z)_{Q1}$	1.82	1.82	1.82	1.82	1.82	1.82	1.82
	$S(Z)_{Q2}$	1.85	1.85	1.85	1.85	1.85	1.85	1.85
	$S(Z)_{Q3}$	1.88	1.88	1.88	1.89	1.88	1.89	1.88
	$S(Z)_{Q4}$	2.42	2.41	2.41	2.44	2.41	2.42	2.41
Roughness	R_{min}	0.026	0.013	0.013	0.043	0.030	0.032	0.027
	R_{ave}	1.65	1.66	1.66	1.90	1.78	1.84	1.74
	R_{Q1}	0.037	0.018	0.018	0.046	0.041	0.043	0.032
	R_{Q2}	0.065	0.035	0.035	0.068	0.064	0.066	0.053
	R_{Q3}	1.94	2.09	2.09	2.82	2.40	2.62	2.31
	R_{Q4}	7.48	7.34	7.34	7.48	7.48	7.48	7.43

On the grounds that users have formed typical ideas about camera shutter sounds, we modified existing shutter sounds and investigated the effects of time structure-related PA. With ten existing cameras, four UX researchers and two sound engineers interpreted qualitative interview data to elicit UV and PA that form a shutter sound's characteristics (envelope shape, total duration, number of peaks, and peak-to-peak duration). From the results of the first jury test, these PAs were controlled or modified to reduce the number of sound samples to eight. Their perceptual characteristics were quantified by psychoacoustic variables (loudness, sharpness A, sharpness Z, and roughness), and subjectively evaluated using a 7-point Likert scale questionnaire for UV.

From the analysis, two user groups, which showed significant differences in age and camera experience, were identified; the respective mean age and camera experience of G1 were 32.0 and 8.4 years, and those of G2 were 28.5 and 5.2 years. The results demonstrated a well-known fact that individual taste influences people when judging sound quality (Pellegrini, 2001; Sköld, Västfjäll, & Kleiner, 2004; Susini et al., 2004). Since we revealed relationships between customer satisfaction and demographic characteristics (age and gender), the results of this study could help to formulate marketing strategies; for example, sound engineers could develop more "soft" (reverse of "hard") shutter sounds for experienced users in their thirties, and develop more "silent" sounds for more inexperienced users in their twenties. However, it should be noted that most of the participants in our study were between the ages of 20 and 40 years. Although we were not able to cover user segments across all possible age groups, the results showed the necessity of specifying user groups depending on the targeted sound impression.

The identified user groups showed the differing influences of PA and psychoacoustic variables on satisfaction level. Since G1's regression model showed significant effects of "hard" and "sophisticated", we examined relative psychoacoustic variables on these UVs; "hard" was negatively related to loudness (N_{ave}, N_{min}, N_{Q3}), sharpness A ($S(A)_{ave}$, $S(A)_{Q3}$), and sharpness Z ($S(Z)_{ave}$, $S(Z)_{Q1}$, $S(Z)_{Q2}$), while "sophisticated" showed no significant relationships. Meanwhile, G2's regression model revealed significant effects of "sophisticated" and "silent". For "sophisticated", G2 participants were significantly influenced by loudness (N_{Q3}) and sharpness A ($S(A)_{Q3}$), and "silent" was significantly related to loudness (N_{ave}, N_{Q3}), sharpness A ($S(A)_{ave}$, $S(A)_{Q3}$), and sharpness Z ($S(Z)_{ave}$). Considering each 130 ms peak-to-peak duration and increasing envelope shapes brings higher values of loudness and sharpness A (Table 4.7), thus it is reasonable to say that their combination affected S3 and S4 sounds for "soft" more for G1, and "sophisticated" and "silent" less for G2.

As to sound samples, significantly higher scores on S4 and S8 were obtained for G1 than G2 ($p < 0.01$). As mentioned above, G2 participants might have not been satisfied by S4, as the composition of the 130 ms peak-to-peak duration and increasing envelope shape induced less "sophisticated" and "silent" feelings. In the meantime, S7 and S8 had a common envelope shape and peak-to-peak duration but showed a substantial difference in satisfaction scores. Only the number of peaks were different in that S8 had 3 peaks while S7 had 4 peaks. From Table 4.7, we were able to infer that 3 peaks showed lower loudness and sharpness A up to the third quartile (N_{min}~N_{Q3} , $S(A)_{min}$~$S(A)_{Q3}$), higher loudness and sharpness A for the fourth quartile (N_{Q4} and $S(A)_{Q4}$), and higher roughness than 4 peaks. Considering higher loudness, sharpness A, sharpness Z, and roughness reduced "hard" and "annoying" feelings (which consequently improved G1's satisfaction scores), higher N_{Q4}, $S(A)_{Q4}$, R_{ave}, and R_{Q3} of S8 might also have improved satisfaction scores. For G2, low- to mid-ranged loudness and sharpness A, rather than the third and the fourth quartiles, influenced satisfaction scores. Therefore, it could be said that the lower values of N_{min}, N_{Q1}, N_{Q2}, $S(A)_{\mathrm{min}}$, $S(A)_{\mathrm{Q1}}$, and $S(A)_{\mathrm{Q2}}$ of S8 reduced "fun" and "classical" feelings, consequentially resulting in lower satisfaction scores of G2 participants.

Throughout this study, we investigated satisfaction level of the participants and were able to reveal the effects of PA on UV by using psychoacoustic variables. This study has several limitations. First, the effect of spectral characteristics was not observed as only time structure-related variables were manipulated. We modified one camera shutter sound to narrow our focus, but when considering its importance in judging sound quality, more research on the effects of spectral characteristics is needed in further research (Shin, Ih, Hashimoto, & Hatano, 2009; Zhang, Huang, Du, & Vertiz, 1996). Second, the results of this study cannot be generalized for other products. Since the perception of sound quality depends on the product categories and context of use (Franinovic & Visell, 2008; Özcan & Jacobs, 2014), it will be necessary to investigate effective UVs for other product sounds.

Despite these limitations, the findings of this study have implications for sound designers and marketing managers. First, we were able to classify user groups by showing significant differences in demographic characteristics. For G1 participants with an average age of 32.0 and 8.4 years of camera experience, satisfaction was positively influenced by "sophisticated," and negatively influenced by "hard". Meanwhile, G2 with an average age of 28.5 and camera experience of 5.2 years showed positive impacts of "sophisticated" and "silent". Second, we analyzed the relationships between PA, psychoacoustic variable, UV, and user satisfaction. Specifically, the two user groups were affected differently by the average and the third quartile of loudness and sharpness A. Although participants in G1 were positively influenced by these psychoacoustic measures, those in G2 were negatively affected. At last, UV, which were elicited from network analysis, were able to explain user's satisfaction scores with adequate goodness-of-fit. Although the result was not different from frequency analysis, characteristics of the collected text data could be observed; for shutter sounds, a network showed a cohesive network structure whose core part was mainly composed of UV. However, the method will be useful in identifying implicit UV for more complicated network. In future research, we hope that the proposed analysis procedure could be applied to investigate other sound impressions and product categories.

CHAPTER 5

Identifying User Values and Product Attributes using Qualitative Data on Vacuum Cleaners

5.1. Overview

This section introduces a novel approach to evaluating the user experience (UX) by calculating the relationship between concepts based on qualitative text data. Classifying concepts into user values (UV) and product attributes (PA) allows the UX quantification model to be suggested by a linear combination of UV, and semantic associations between UV and PA were identified.

Since it is highly subjective and context-dependent, only marginal success has been achieved in the quantification of UX. Existing studies have defined and quantified UX dimensions and analyzed their relationships by utilizing statistical analysis methods. A multiple regression model (Asche & Kreis, 2014), structural equation model (Knijnenburg et al., 2012; Park et al., 2014), and nonlinear models were applied and compared by Park et al. (2013). However, those studies had limitations that their numerical values should be collected from users by applying multiple user research methodologies such as questionnaires (Adikari, McDonald, & Campbell, 2011), physiological measures (Ganglbauer, Schrammel, Deutsch, & Tscheligi, 2009), and observational techniques (Hassenzahl & Tractinsky, 2006). This section overcomes this limitation by suggesting a method to develop a UX quantification model based on text data.

The importance of keywords in qualitative data has been recognized by the occurrence and co-occurrence of keywords. Although tf–idf (term frequency–inverse document frequency) is one of the most representative methods with which to represent a term's importance, it considers rare terms more important when detecting differences between documents, which is not appropriate for our research purpose. Therefore, the centrality measures of a semantic network and simple term frequency were applied to weigh keywords (Salton & Buckley, 1988). Table 5.1 shows the literature in which centrality measures were applied. However, such literature utilized the methods of summarizing or categorizing documents rather than identifying the UX.

After identifying the UVs that compose UX, the relationships between PA were also extracted. UX researchers make many decisions while developing products and services. They should be aware of the relationship between UV and PA to make better decisions. Several methodologies that have been used to identify relationships such as Quality Function Deployment (QFD) (Iranmanesh, Thomson, & Salimi, 2005), Interpretive Structural Modeling (ISM) (Warfield, 1973), and Analytical Hierarchy Process (AHP) (Saaty, 1990) were suggested. Those methods help understand the relationships between components in a complex system, but are usually based on a predefined hierarchical structure or the properties of a product, which requires either experts' explicit knowledge or numerical scores for PA.

Table 5.1. Literatures using centrality measures in weighting keywords

Author (Year)	Title	Measure	Summary
Iezzi (2012)	Centrality Measures for Text Clustering	degree, betweenness, closeness, eigenvector	A new method to filter terms without ignoring the context was introduced.
Oya (2013)	Degree centralities, closeness centralities, and dependency distances of different genres of texts	degree, closeness	Distribution of degree and closeness centralities are affected by the number of sentences and genres of corpus.
Abilhoa & De Castro (2014)	A keyword extraction method from twitter messages represented as graphs	degree, closeness	Network analysis was compared to tf-idf and KEA algorithm, and showed a superior performance in extracting important keywords.
He & Tan (2015)	Study on SINA micro-blog personalized recommendation based on semantic network	degree	User interest topic was extracted by adopting semantic network analysis on Micro-blog.
Hong, Shin, & Kim (2016)	High/low reputation companies' dialogic communication activities and semantic networks on Facebook: A comparative study	degree, betweenness, closeness, eigenvector	Facebook posts and comments of five corporates were compared to reveal differences in reputation and communication.
Kang et al. (2017)	Semantic network analysis of vaccine sentiment in online social media	degree, betweenness, closeness, eigenvector	Positive, neutral, and negative topics on vaccine were found by calculating centralities of semantic networks.

For example, the three PA of hardness, consistency, and thickness for automobile panels were identified by domain experts by conducting conjoint analysis (Rhiu, Ryu, Jin, & Yun, 2011). Meanwhile, the concept of participatory design, which involves general users during the product development phase, was adopted. However, the method presents specific questions such as, "*What kind of product Y should we create for people X? What Y will fit the needs and preferences of X?*" While this method helps specify customer preferences, it is both difficult to survey a large number of participants, but this also takes much time and effort from designers (Pals, Steen, Langley, & Kort, 2008).

Statistical analysis without theoretical interpretation may lead to a spurious correlation problem. As implied by the famous phrase "correlation does not imply causation," a significant relationship between PA and UV does not always imply the effectiveness of PA on UV (Aldrich, 1995; Jacobs, Leamer, & Ward, 1979; Shipley, 1999). In their research, Tufte (2003) mentioned "empirically observed covariation is a necessary but not sufficient condition for causality," emphasizing the necessity of investigating the association between PA and UV in addition to statistical analysis. There were also attempts to extract PA from text descriptions based on word frequency and co-occurrence, but this did not relate them to specific UV (Popescu & Etzioni, 2007; Raju, Shishtla, & Varma, 2009; Scaffidi et al., 2007).

As a case study, interview transcripts for vacuum cleaners with two tasks (pull & push and storage) were analyzed. This chapter introduces two issues of (1) suggesting a UX quantification model, and (2) identifying the relationship between UV and PA. First, the importance of UV on vacuum cleaner choice was elicited, and UX quantification models were proposed using seven centrality measures as UV weights. The prediction powers of the suggested models were compared to the quantitative study and thus revealed fairly high levels for the coefficient of determination. Second, the relationship between UV and PA were identified based on a simple hypothesis: if UV and PA are related, terms will co-occur. We generated a subnetwork on each PA, and calculated the centralities of UV for each subnetwork. The result was used to assist the result of correlation analysis; since network analysis involves a semantic level of association between two concepts, the method will be able to help avoid spurious correlations in quantitative research.

5.2. Method

The research process in this section is organized into three phases (Figure 5.1): (1) Elicit UV, (2) Suggest a UX quantification model, and (3) Identify UV that are relevant to PA.

First, text data was collected and preprocessed to form a co-occurrence matrix. Two qualitative research studies on vacuum cleaners were introduced as Datasets 1 and 2. The user expression data in the context of pulling and pushing the vacuum cleaners were collected for Dataset 1, and descriptions in the context of storing the vacuum cleaners were collected as Dataset 2. Participants were directed to express their feelings in detail while using three different vacuum cleaners and semantic network analysis was conducted based on the co-occurrence matrix; the node-level centralities of keywords were calculated and arranged in descending order. For case studies in this paper, the top 10 keywords of network centralities (degree, closeness, betweenness, and eigenvector centralities) were used to elicit important UVs.

Second, UX was expressed in form of linear equations and denoted the "UX quantification model." During this process, seven different network centralities were used to give weights to UV, which were elicited from the former process. The goodness-of-fit of the proposed models were compared to the results of the quantitative study to validate the suggested research method's effectiveness.

Third, the relationship between UV and PA were identified. Important PAs were elicited from the top 10 keywords of network centralities, and their relationships with UVs were examined by generating PA-subnetworks. Since the co-occurrence of terms represented the association in a semantic space, we used the result as a subsidiary of the correlation analysis.

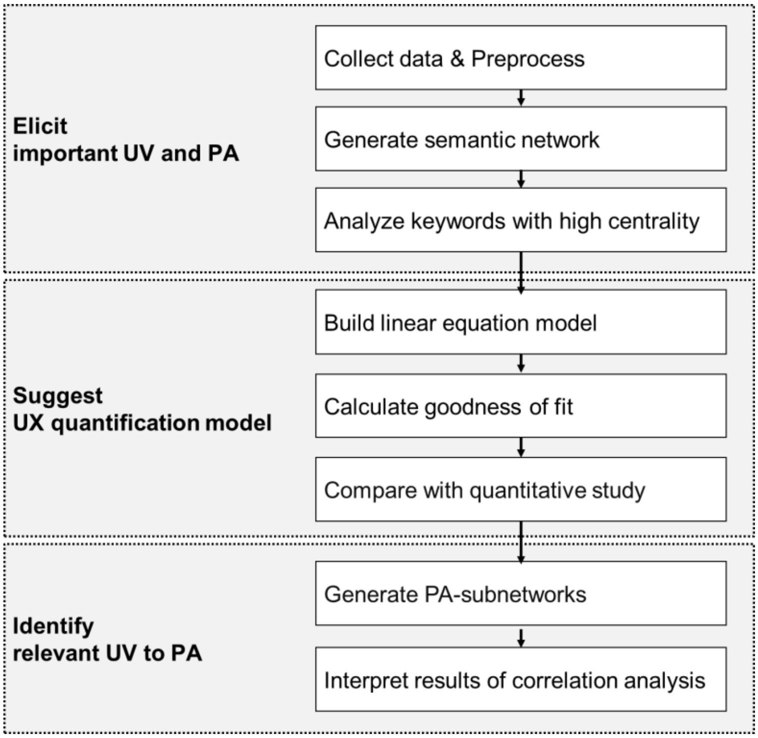


Figure 5.1. Research process for identifying UV and PA

5.2.1. Eliciting important UV and PA

5.2.1.1. Collect and preprocess data

As case studies, qualitative data were collected using think-aloud methods while using vacuum cleaners. Users freely expressed their feelings while pulling and pushing and while storing vacuum cleaners. In total, 21 housewives with an average age of 48.55 were recruited, and they were instructed to use three types of vacuum cleaner in a random sequence.

Dataset 1 collected users' verbal expressions while using the vacuum cleaners. This task will be referred as the "push and pull" task in the rest of this paper. The participants were instructed to clean coffee powder that had intentionally been distributed on the floor. The nature of the coffee powder distribution required that participants use vacuum cleaners not only on the floor, but also under various types of furniture (a desk, bookshelf, and couch).

Dataset 2 collected users' thoughts when storing the vacuum cleaners. After conducting the pull and push task, participants were instructed to store the vacuum cleaner that they had used. Participants freely expressed their feelings and thoughts during and after storing them. Verbally expressed words were typed and used in further analysis.

Figure 5.2. Experimental environments of Dataset 1 (left) and Dataset 2 (right)

A preprocessing step should be conducted before generating a semantic network. This step includes defining word boundaries, stop words, and synonyms. We used an automation tool developed by Lee et al. (2014) that helps extract a co-occurrence matrix from unstructured natural language data.

First, the scope in which co-occurrence can take place, i.e. word boundaries, was defined. For case studies in this dissertation, the terms that come out from the same participants were considered to have co-occurred. Second, the word stemming step that decomposes clauses into morphemes was proceeded. The Lucene Korean analyzer (http://cafe.naver.com/korlucene) that is mostly used to analyze Korean was utilized. Third, stop words such as *a*, *the*, and *to* and other designated words were eliminated. In addition, synonyms such as *year* and *yrs* were unified into the same form. When determining synonyms, morphemes were grouped into a smaller number of keywords to prevent the multicollinearity problem in the UX quantification model. In this process, five UX researchers shared opinions to prevent the involvement of subjectivity.

5.2.1.2. Analyze semantic network

Co-occurring keywords were encoded into an edge between two nodes where each node represented a keyword. Node and edge information was used to obtain a relational matrix with which to calculate node-level statistics including the degree, closeness, betweenness, and eigenvector centralities by utilizing UCINET 6 (Borgatti et al., 2002). The most popular network centralities, which are degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality, were calculated and observed. The normalized centrality measures of the collected terms are listed in Appendix B.

5.2.1.3. Analyze keywords with high centrality

The keywords in a network were classified into three categories: UV, product attributes (PA), and contextual variable (CV). UV indicates the adjectives and adverbs that contain users' subjective feelings or emotions. PA indicates the physical parts and properties of a product that can be measured, such as the *handle* and *handle length*. CV encompasses the perceptual expressions of the objects, such as its *looks* and *grip*, environments such as *furniture* and *storage*, and participants' behaviors such as *push* and *rolling*. Categories of keywords were examined by matching with the following sentences: "(CV) of (PA) is (UV)," or "(PA) is (UV) when (CV)". For example, "the *design* (CV) of a *handle* (PA) is *comfortable* (UV)." or "the *head* (PA) is *tight* (UV) when cleaning under the *furniture.*"

For the case studies, keywords with the top 10 centralities were selected to evaluate UVs. PAs were not considered in this step, and UV and CV were interpreted to determine the UV of each dataset.

5.2.2. Suggesting UX quantification model

5.2.2.1. Build linear equation model

After eliciting the UV, UX quantification models were suggested. The equations were presented in a linear equation form using UVs as independent variables to predict user satisfaction. The logic of the UX quantification model is straightforward: the higher the UV centrality, the greater the influence of the UV on overall user satisfaction.

The weights of UV were calculated by the normalized centrality measures. Seven different centrality measures (degree centrality, Bonacich power, Freeman betweenness centrality, flow betweenness centrality, eigenvector centrality, closeness centrality, and ARD) were used to build the models. The degree centrality, Bonacich power, and eigenvector centrality are frequency-based measures, the closeness centrality and ARD are distance-based measures, and the Freeman betweenness centrality and flow betweenness centrality are path-based measures. In addition to the network centrality, term frequency was used to weigh UVs, since frequency was the most basic method with which to determine a word's importance.

5.2.2.2. Calculating the goodness of fit

The fitness of UX quantification models were examined by the coefficient of determination (R^2). The R^2 value indicates the proportion of the variance in the dependent variable that is predictable from the independent variables, as calculated by the equation $R^2 = S_{xy}^2 / S_{xx} \cdot S_{yy}$. In addition, the Variance Inflation Factor (VIF) was calculated to examine the multicollinearity of variables. The VIF of the *i*-th predictor was obtained by regressing the variable (x_i) on the remaining predictors (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n). After obtaining the R^2 of *i*-th predictor (R_i^2), the VIF of *i*-th predictor was calculated by $\mathrm{VIF}_i = 1/(1-R_i^2)$.

5.2.2.3. Comparison with quantitative study

The goodness-of-fit of the UX quantification models were compared to the results of the quantitative study. Numerical scores on UV were collected by conducting a usability test; participants conducted the same tasks on vacuum cleaners and rated the numerical scores on a 7-point Likert scale. The questionnaire items were composed of UVs, which were elicited from network analysis.

Usability tests of vacuum cleaners were conducted with 42 participants composed of 12 males and 30 females with average age 38.0 years. For the pull and push task that corresponds to Dataset 1, participants used six different vacuum cleaners in a random order, and vacuumed coffee powder on the floor, carpet, and under furniture. In the case of the storage task, which corresponds to Dataset 2, participants arranged all six vacuum cleaners before evaluating each sample. They were allowed to compare the vacuum cleaners while filling out the questionnaires. Numerical scores on UV were collected based on 7-point Likert scale, and satisfaction scores were based on a 100-point scale.

The collected data were analyzed via linear regression. Stepwise regression estimates the parameters of independent variables using the least squares method, while automatically excluding ineffective variables. As a result of regression analysis, R^2 and the VIF of regression models were also calculated. The analysis was conducted using SPSS for Windows version 16.0.

5.2.3. Identifying relevant UV to PA

5.2.3.1. Generate PA-subnetworks

Based on Datasets 1 and 2, we examined the relationship between UV and PA based on a simple hypothesis: if UV and PA are associated, the terms will co-occur. With this basic assumption, we generated a subnetwork of each PA and calculated the centralities of each subnetwork.

First, we selected the PAs that appear for the top 10 centralities (degree, closeness, eigenvector, and the betweenness centrality). Unlike building the UX quantification model in which only UVs were used, we also listed PAs from the top 10 keywords. As we had assumed that related UVs and PAs co-occurred, we generated a subnetwork by screening sentences that contained a certain PA. After filtering phrases that contained a certain PA, co-occurrence matrices were formed. As an example, to generate a "head" subnetwork, phrases that contained the word "head" were selected to build a co-occurrence matrix. The subnetworks generated from these co-occurrence matrices are called PA-subnetworks in this dissertation.

5.2.3.2. Correlation analysis between PA and UV

While qualitative research was conducted with three vacuum cleaners, the quantitative research used six different vacuum cleaners, as shown in Figure 5.3. The design specifications such as the weight, width, length, size, and noise for six vacuum cleaners were measured and listed in Table 5.2. Here, the width and length of the "hose" represented the overall width and length of the vacuum cleaners including the hose, and the weight of the "hose" included the weight of the "handle," as those were not detached during the experiment.

The relationships between UV and PA were examined via correlation analysis. Since only six samples were observed, Spearman's rank correlation coefficient was conducted, which is useful for assessing the monotonic relationships between two variables. The method of calculating Spearman's rank correlation coefficient is as follows. If there is a dataset with a sample size n and ranks X_i and Y_i are denoted as $\mathrm{rg}X_i$ and $\mathrm{rg}Y_i$, the correlation coefficient is calculated with the equation $\rho\ (\mathrm{rg}X_i\ ,\ \mathrm{rg}Y_i) = \mathrm{cov}\ (\mathrm{rg}X_i\ ,\ \mathrm{rg}Y_i)\ /\ \sigma(\mathrm{rg}X_i)\ \sigma(\mathrm{rg}Y_i)$.

Table 5.2. The design specification of vacuum cleaners

		S1	S2	S3	S4	S5	S6
Head	weight (kg)	0.24	0.72	0.68	0.58	0.34	0.4
	thickness (mm)	23	41	40	27	15	23
	width (mm)	310	300	286	308	259	310
	length (mm)	78	118	118	104	79	82
	size (m^2)	0.02	0.04	0.03	0.03	0.02	0.03
	head noise (dB)	55	67	56	57	66	74
Body	weight (kg)	5.64	7.52	6.02	5.96	4.48	6.94
	width (mm)	300	320	430	300	250	430
	length (mm)	540	680	300	690	630	530
	volume(m^3)	0.17	0.24	0.13	0.16	0.14	0.23
	wheel noise (dB)	73	73	65	75	78	77
Hose + Body	width (mm)	300	460	430	300	370	580
	length (mm)	640	840	300	760	690	720
	area (m^2)	0.19	0.39	0.13	0.23	0.26	0.42
Stick	weight (kg)	0.54	0.92	0.84	0.54	0.68	0.96
	angle (°)	82.7	80	90.1	43	74.6	81.9
Handle	weight (kg)	0.24	0.66	0.18	0.54	0.18	0.66
	wrist angle (°)	6.4	28.4	11.1	6.5	4.8	5.7

Figure 5.3. Product samples used in quantitative research

5.3. Results

5.3.1. Important UVs

From the generated semantic networks, centralities of each dataset were calculated and listed in a descending order (Tables 5.3 and 5.4). Degree centrality, Freeman betweenness centrality, closeness centrality, and eigenvector centrality were used to select the important keywords. A list of all keywords and centralities are shown in Appendix B. A semantic network of Datasets 1 and 2 are illustrated in Figures 5.4 and 5.5 and the sizes of the nodes in the figures indicate the eigenvector centralities in the network, the dark nodes with a round shape indicate UV, gray nodes with a square shape indicate PA, and the bright nodes with a diamond shape indicate CV. Thicker links indicate stronger connections.

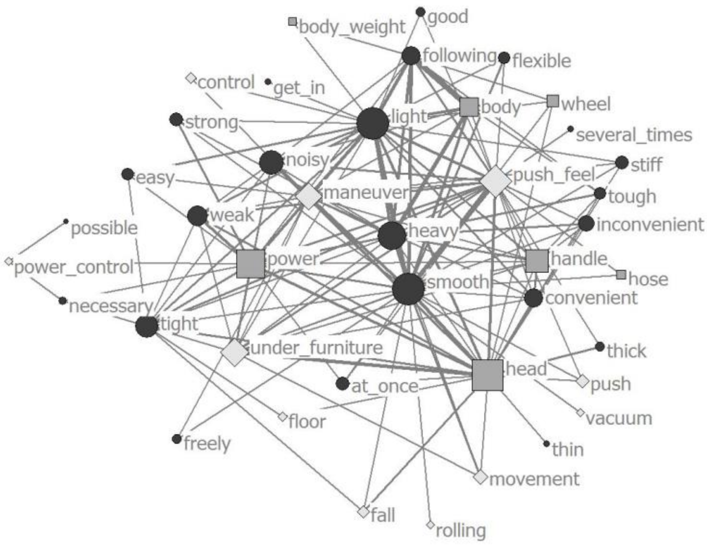


Figure 5.4. The semantic network of Dataset 1

Note: The sizes of the nodes indicate their eigenvector centralities in the network. The dark nodes with a round shape are UV, the gray nodes with a square shape are PA, and the bright nodes with a diamond shape are CV.

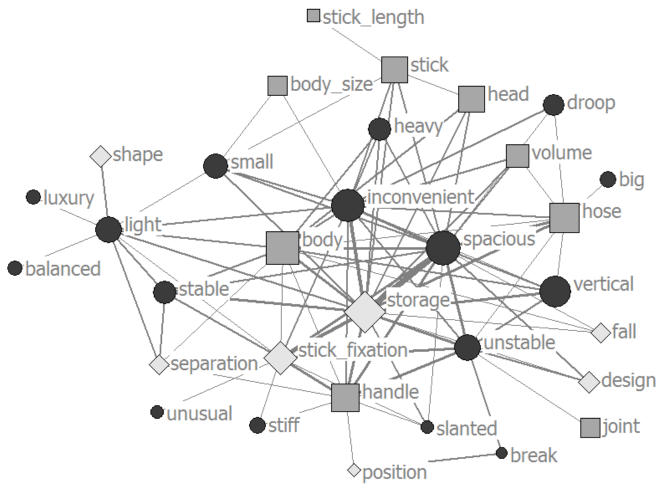


Figure 5.4. The semantic network of Dataset 2

Note: The sizes of the nodes indicate their eigenvector centralities in the network. The dark nodes with a round shape are UV, the gray nodes with a square shape are PA, and the bright nodes with a diamond shape are CV.

The network centralities of Dataset 1 are listed in Table 5.3. Among these, terms about the power of vacuum cleaners were not considered (*strong* and *weak*) to limit our focus to the pulling and pushing task. In addition, CVs were analyzed to extract implicit user needs; *Push feel* was considered a higher concept than other UV such as *smooth*, *light*, and *heavy*. Meanwhile, *under furniture* was replaced by the term *deep*, as it implied a user's need to reach deeper places under furniture. Finally, *maneuver* was simply replaced with *maneuverable*. Finally, UV for the pulling and pushing task were evaluated with seven keywords: "smooth," "deep," "light," "maneuverable," "following," "convenient," and "quiet."

Table 5.3. Top 10 centralities for vacuum cleaner pull & push (*Dataset 1*)

Degree		Closeness		Eigenvector		Betweenness	
head^{P}	0.695	head^{P}	0.753	head^{P}	0.316	head^{P}	0.133
push feel^{C}	0.641	push feel^{C}	0.724	push feel^{C}	0.309	push feel^{C}	0.087
smooth^{U}	0.603	smooth^{U}	0.704	smooth^{U}	0.299	smooth^{U}	0.071
under furniture^{C}	0.580	under furniture^{C}	0.693	under furniture^{C}	0.293	following^{U}	0.067
light^{U}	0.519	light^{U}	0.665	power^{P}	0.279	body^{P}	0.065
power^{P}	0.496	power^{P}	0.652	light^{U}	0.277	under furniture^{C}	0.065
noisy^{U}	0.450	noisy^{U}	0.636	heavy^{U}	0.25	light^{U}	0.049
heavy^{U}	0.450	heavy^{U}	0.636	maneuver^{C}	0.248	noisy^{U}	0.047
handle^{P}	0.450	handle^{P}	0.636	weak^{U}	0.241	handle^{P}	0.041
weak^{U}	0.427	inconvenient^{U}	0.624	noisy^{U}	0.239	heavy^{U}	0.034

Note: Superscripts indicate keyword categories: U indicates UV, P indicates PA, and C indicates CV.

For Dataset 2, UVs on vacuum cleaner storage were elicited from Table 5.4. Here, *spacious* refers to the vacuum cleaner's compactness, so users were not disturbed during their everyday lives. UV from *vertical* was able to be replaced with *spacious*, as users tried to save space through standing the body of the vacuum cleaners. *Stick fixation* was represented by *easy fixation*, as the majority of users wanted to place the stick more easily. Meanwhile, *design* was interpreted as related to the term *sophisticated*. Therefore, the five UV of "spacious," "easy fixation," "unstable," "light," and "sophisticated" remained.

Table 5.4. Top 10 centralities for vacuum cleaner storage (*Dataset 2*)

Degree		Closeness		Eigenvector		Betweenness	
inconvenient^{U}	0.553	inconvenient^{U}	0.673	spacious^{U}	0.375	inconvenient^{U}	0.144
spacious^{U}	0.526	spacious^{U}	0.673	inconvenient^{U}	0.351	handle^{P}	0.133
handle^{P}	0.474	handle^{P}	0.650	body^{P}	0.351	vertical^{C}	0.078
body^{P}	0.474	body^{P}	0.650	vertical^{C}	0.321	spacious^{U}	0.072
vertical^{C}	0.447	vertical^{C}	0.633	stick fixation^{C}	0.316	stick fixation^{C}	0.065
stick fixation^{C}	0.421	stick fixation^{C}	0.628	hose^{P}	0.297	body^{P}	0.058
hose^{P}	0.368	hose^{P}	0.613	handle^{P}	0.279	stable^{U}	0.057
unstable^{U}	0.355	unstable^{U}	0.598	light^{U}	0.269	unstable^{U}	0.04
light^{U}	0.303	light^{U}	0.576	stick^{P}	0.258	hose^{P}	0.034
stable^{U}	0.303	stable^{U}	0.576	unstable^{U}	0.255	design^{C}	0.029

Note: Superscripts indicate keyword categories: U indicates UV, P indicates PA, and C indicates CV.

5.3.2. UX quantification model

A weighted sum of UV was calculated to predict the overall satisfaction level and build a linear equation model. The weight of each UV was drawn from the centrality measures listed in Tables 5.5 and 5.6. Since we assumed that the centrality measures were proportional to the effects of the keywords on customer satisfaction, different equations were developed for different centralities. In addition, min–max normalization values of term frequency were also used to weigh UV.

Furthermore, usability tests were conducted that corresponded to the interview data. Numerical scores on UV were collected on 7-point Likert scales, and satisfaction scores were collected on 100-point scales. The fitness of the UX quantification model was verified via R^2 and VIF.

5.3.2.1. UX quantification model of datasets

Tables 5.5 and 5.6 respectively show the normalized values of network centralities on UV for Datasets 1 and 2. Seven different network centralities and term frequencies were used to build the UX quantification models. In the case of using the degree centrality as UV weights, the user satisfaction level while pulling and pushing was presented as in Equation (5.1), and the user satisfaction level during the storing task was presented by Equation (5.2).

$$\begin{aligned}\text{Satisfaction (Pull \& Push)} = {} & 0.603 \times \text{smooth} + 0.580 \times \text{deep} + \\ & 0.519 \times \text{light} + 0.450 \times \text{quiet} + 0.420 \times \text{maneuverable} + \\ & 0.397 \times \text{following} + 0.328 \times \text{convenient}\end{aligned} \tag{5.1}$$

$$\begin{aligned}\text{Satisfaction (Storage)} = {} & 0.526 \times \text{spacious} + 0.421 \times \text{easy fixation} + \\ & 0.355 \times \text{stable} + 0.303 \times \text{light} + 0.184 \times \text{sophisticated}\end{aligned} \tag{5.2}$$

Table 5.5. Parameters of UX quantification model for Dataset 1

	smooth	deep	light	quiet	following	conve-nient	maneu-verable
Degree	0.603	0.580	0.519	0.45	0.397	0.328	0.420
BonPow	0.945	0.924	0.874	0.753	0.632	0.644	0.784
Between	0.071	0.065	0.049	0.047	0.067	0.011	0.018
FlowBet	0.724	0.515	0.628	0.514	1.000	0.377	0.258
Eigenvec	0.299	0.293	0.277	0.239	0.2	0.204	0.248
Closeness	0.704	0.693	0.665	0.636	0.621	0.59	0.621
ARD	0.914	0.893	0.836	0.771	0.726	0.657	0.740
tf	60	24	50	22	55	16	17

Note: "BonPow" indicates Bonacich power, "Between" indicates betweenness, "FlowBet" indicates flow betweenness, and "Eigen" indicates eigenvector, "tf" indicates term frequency.

Table 5.6. Parameters of UX quantification model for Dataset 2

	spacious	easy fixation	stable	light	sophisticated
Degree	0.526	0.421	0.355	0.303	0.184
BonPow	0.872	0.732	0.587	0.621	0.365
Between	0.072	0.065	0.040	0.009	0.029
FlowBet	0.621	0.338	0.543	0.392	0.171
Eigen	0.375	0.316	0.255	0.269	0.162
Closeness	0.673	0.628	0.598	0.576	0.543
ARD	0.831	0.850	0.643	0.582	0.469
tf	26	25	15	6	4

Note: "BonPow" indicates Bonacich power, "Between" indicates betweenness, "FlowBet" indicates flow betweenness, and "Eigen" indicates eigenvector, "tf" indicates term frequency.

5.3.2.2. Goodness-of-fit of UX quantification models

The fitness of the suggested UX quantification models were examined by calculating the R^2 and VIF values. The results of Datasets 1 and 2 are as listed in Tables 5.7 and 5.8. For Dataset 1, the Bonacich power, eigenvector centrality, and closeness centrality showed the highest R^2 value, while Dataset 2 showed the highest R^2 value for the closeness centrality. Nevertheless, there was not much difference in R^2 values between different network centralities. In addition, no multicollinearity problem appeared, thus values under 10 were considered acceptable (Kupper, 1978).

Table 5.7. Goodness-of-fit of UX quantification for Dataset 1

	R^2	Max. VIF
Degree	0.511	1.942
Bonacich Power	**0.523**	**1.989**
Betweenness	0.438	1.694
Flow Betweenness	0.458	1.785
Eigenvector	**0.523**	**1.989**
Closeness	**0.523**	**1.992**
ARD	0.519	1.976
Term Frequency	0.451	1.747

Table 5.8. Goodness-of-fit of UX quantification for Dataset 2

	R^2	Max. VIF
Degree	0.461	1.958
Bonacich Power	0.456	1.994
Betweenness	0.480	1.885
Flow Betweenness	0.465	2.024
Eigenvector	0.457	2.002
Closeness	**0.487**	**2.142**
ARD	0.479	2.068
Term Frequency	0.444	1.878

5.3.2.3. Linear regression analysis

As a result of quantitative research, numerical scores for UV were collected and analyzed by the stepwise regression method. As in the UX quantification model, the models' fitness was examined by R^2 and VIF. The results for each dataset are as shown in Tables 5.9 and 5.10. Unlike the UX quantification model, the regression coefficients of UV were examined to remove insignificant variables. In this step, "following" and "quiet" were insignificant for the pulling and pushing task, as was "light" for the storing task. The variables might have been removed by the limitation of the product samples.

The results of linear regression analysis showed slightly higher performances than the UX quantification models. The UX quantification models showed the highest R^2 value of 0.523 (Bonacich power, eigenvector centrality, and closeness centrality), for Dataset 1, and 0.487 (closeness centrality) for Dataset 2, while those of the regression models were 0.604 and 0.542.

Table 5.9. Results of stepwise linear regression analysis for Dataset 1

R^2 = 0.604, Max.VIF=1.543

	unstandardized		standardized	t	sig.	VIF
	B	std. Error	Beta			
(constant)	-0.022	0.034		-0.637	0.524	
maneuver	0.452	0.048	0.480	9.389	<0.001	1.543
smooth	0.107	0.045	0.119	2.396	0.017	1.460
convenient	0.145	0.041	0.158	3.550	<0.001	1.176
light	0.161	0.047	0.176	3.447	<0.001	1.539
deep	0.126	0.043	0.134	2.947	0.004	1.219

Table 5.10. Results of stepwise linear regression analysis for Dataset 2

R^2 = 0.542, Max.VIF = 1.571

	unstandardized		standardized	t	sig.	VIF
	B	std. Error	Beta			
(constant)	0.002	0.035		0.046	0.964	
spacious	0.195	0.053	0.204	3.679	<0.001	1.571
easy fixation	0.151	0.051	0.157	2.938	0.004	1.460
stable	0.387	0.046	0.404	8.369	<0.001	1.196
sophisticated	0.265	0.046	0.282	5.690	<0.001	1.258

5.3.3. Relationship between UV and PA

For the PA-subnetworks of Dataset 1, 97, 65, and 60 keywords were extracted in total from the "head," "body," and "handle" subnetworks, respectively. For Dataset 2, 43, 41, 42, and 30 keywords were left for "handle," "body," "stick," and "hose" subnetworks, respectively. The generated subnetworks were as shown in Figures 5.6 and 5.7.

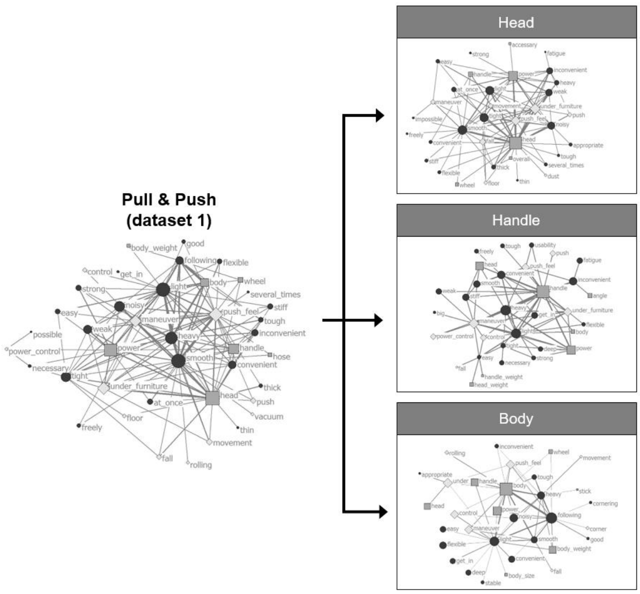


Figure 5.6. Subnetwork generation of Dataset 1

Note: Nodes with link length over two are represented and the sizes of the nodes indicate the eigenvector centralities in the network. The dark nodes with a round shape are UV, the gray nodes with a square shape are PA, and the bright nodes with a diamond shape are CV. Thicker links indicate stronger connections.

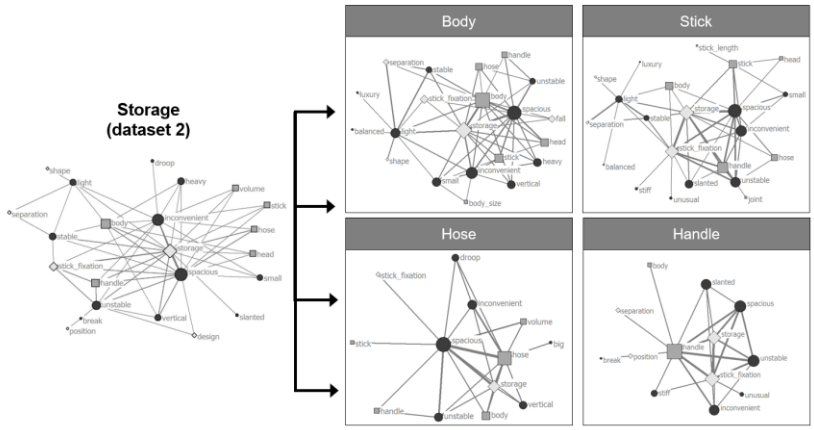


Figure 5.7. Subnetwork generation for Dataset 2

Nodes with link length over four are represented. The sizes of the nodes indicate the eigenvector centralities in the network. The dark nodes with a round shape are UV, the gray nodes with a square shape are PA, and the bright nodes with a diamond shape are CV. Thicker links indicate stronger connections.

5.3.3.1. Network centrality of the PA-subnetworks

The relationships between UV and PA were evaluated based on the assumption that relevant UV and PA will co-occur. Tables 5.11 and 5.12 represent the number of co-occurrences, i.e. the link strength between UV and PA. The weakest link strength between UV and PA was 21 for Dataset 1 and 14 for Dataset 2.

In Dataset 1, the centrality measures of seven UV (“maneuverable,” “smooth,” “following,” “convenient,” “deep,” “light,” and “quiet”) were evaluated for the “body,” “handle,” and “head” subnetworks. The result indicates that “handle” was not related to “following” and “quiet.” Except for these variables, the weakest link strength appeared between “convenient” and “body,” with 21 co-occurrences.

Table 5.11. The centrality measures of PA-subnetworks for Dataset 1

		maneu-verable	smooth	follow-ing	conve-nient	deep	light	quiet
link strength	body	31	89	203	21	47	136	51
	handle	95	63	-	75	65	101	-
	head	78	183	25	98	188	138	78
degree	body	0.266	0.313	0.703	0.234	0.391	0.422	0.359
	handle	0.695	0.559	-	0.525	0.492	0.644	-
	head	0.396	0.625	0.135	0.490	0.719	0.510	0.313
closeness	body	0.561	0.582	0.762	0.561	0.610	0.627	0.587
	handle	0.756	0.686	-	0.670	0.663	0.728	-
	head	0.619	0.722	0.533	0.658	0.774	0.667	0.589
eigenvector	body	0.265	0.244	0.427	0.225	0.328	0.347	0.302
	handle	0.364	0.294	-	0.298	0.276	0.346	-
	head	0.228	0.319	0.108	0.266	0.348	0.279	0.176
betweenness	body	0.004	0.020	0.200	0.006	0.035	0.038	0.020
	handle	0.060	0.038	-	0.025	0.023	0.046	-
	head	0.019	0.060	0.000	0.038	0.091	0.038	0.009

The UV of Dataset 2 were “spacious,” “easy fixation,” “stable,” “light,” and “sophisticated.” The centrality measures of these UV for the “body,” “hose,” “stick,” and “handle” subnetworks are presented in Table 5.12, indicating that “handle” was not related to “light.” The weakest UV–PA link strength appeared for the term, “sophisticated,” with 14 link strength for “body,” “hose,” “stick,” and “handle.”

Table 5.12. The centrality measures of PA-subnetworks for Dataset 2

		spacious	easy fixation	stable	light	sophisticated
link strength	body	113	36	35	65	14
	hose	80	24	35	17	14
	stick	111	99	69	48	14
	handle	57	62	63	-	14
degree	body	0.750	0.550	0.400	0.575	0.275
	hose	0.857	0.571	0.571	0.321	0.393
	stick	0.714	0.762	0.500	0.476	0.262
	handle	0.390	0.390	0.488	-	0.268
closeness	body	0.784	0.690	0.615	0.690	0.580
	hose	0.848	0.700	0.683	0.583	0.622
	stick	0.764	0.792	0.656	0.646	0.568
	handle	0.612	0.612	0.651	-	0.577
eigenvector	body	0.390	0.284	0.242	0.315	0.171
	hose	0.429	0.321	0.338	0.181	0.255
	stick	0.419	0.390	0.294	0.287	0.189
	handle	0.369	0.369	0.404	-	0.282
betweenness	body	0.063	0.029	0.005	0.028	0.000
	hose	0.083	0.017	0.011	0.000	0.000
	stick	0.073	0.176	0.030	0.031	0.000
	handle	0.008	0.008	0.043	-	0.000

5.3.3.2. The result of correlation analysis

The results of quantitative research were as shown in Tables 5.13 and 5.14. The results showed that UV and PA had significant co-occurrence relationships. However, although “following” and “quiet” did not appear in handle-subnetwork, a low but significant relationship was found in Table 5.13.

For Dataset 1, The PAs with the highest correlation coefficient were as follows: “Maneuver” was affected by the weight of “head,” the wrist angle of “handle,” and the length of “body”; “Smooth” was influenced by the weight of “head,” the wrist angle of “handle,” and the weight of “body”; “Following” was affected by the area of “head” and the weight of “body”; “Convenient” was affected by weight of “head,” the wrist angle of “handle,” and the length of “body”; “Deep” was affected by length of “head,” the wrist angle of “handle,” and the weight of “body”; “Light” was influenced by the weight of “head,” the wrist angle of “handle,” and the weight of “body”; “Quiet” was weakly related to the weight of “head,” and the sound pressure of motor noise from “body.”

In the storage task, the PA with the highest correlation coefficients were as follows: “Spacious” was effected by the width of “body,” the width and weight of “hose,” the angle of “stick,” and the weight of “handle”; “Easy fixation” was affected by the length of both “body” and “hose” and the angle of “stick”; “Stable” was weakly related to the volume of “body,” the width and weight of “hose,” the angle of “stick,” and the weight of “handle”; “Light” was affected by the weight of “body,” the width and weight of “hose,” and the weight of “stick”; “Sophisticated” was weakly affected by the length of “body,” the width of “hose,” and the angle of “stick.” The result shows that “sophisticated” was affected by the overall design, rather than some specific part of the vacuum cleaners.

Table 5.13. Result of **c**orrelation analysis of PA and UV for pulling and pushing vacuum cleaners

		maneuverable	smooth	following	convenient	deep	light	quiet
Head	weight (kg)	-0.291**	-0.360**	-0.316**	-0.311**	-0.531**	-0.448**	-0.128*
	thickness (mm)	-0.233**	-0.336**	-0.336**	-0.274**	-0.470**	-0.430**	-
	width (mm)	-	0.167*	-	-	0.230**	-	-
	length (mm)	-0.269**	-0.354**	-0.306**	-0.280**	-0.533**	-0.435**	-
	size (mm^2)	-0.253**	-0.342**	-0.337**	-0.288**	-0.495**	-0.441**	-
	head noise (dB)	-	0.227**	0.246**	-	0.303**	0.260**	-
Handle	weight (kg)	-0.141*	-0.210**	(-0.168**)	-0.102**	-0.355**	-0.218**	-
	wrist angle (°)	-0.277**	-0.275**	(-0.259**)	-0.293**	-0.430**	-0.362**	(-0.157*)
Body	weight (kg)	-0.226**	-0.285**	-0.297**	-0.245**	-0.436**	-0.369**	-
	wheel noise (dB)	-0.180**	-	-	-0.168**	-0.171**	-	-0.139*
	width (mm)	-	-0.206**	-0.219**	-	-0.362**	-0.250**	-
	length (mm)	-0.243**	-	-	-0.301**	-	-0.182**	-0.163*

Note: If no relationship was found in network analysis, correlation coefficients are denoted by a round bracket.
* $p < 0.05$, ** $p < 0.01$.

Table 5.14. Result of **c**orrelation analysis of PA and UV for storing vacuum cleaners

		spacious	easy fixation	stable	light	sophisticated
Body	weight (kg)	0.190**	-	-	-0.448**	-
	volume (m^3)	-	-	0.139*	-	-
	width (mm)	0.305**	0.141*	-	-0.362**	-
	length (mm)	-0.345**	-0.312**	-0.192**	-	-0.138*
Hose	width (mm)	0.255**	-	0.147*	-0.270**	0.128*
	length (mm)	-0.207**	-0.308**	-	-0.207**	-
	area (m^2)	-	-	-0.148*	-	-
	weight (kg)	0.255**	-	0.147*	-0.270**	-
Stick	weight (kg)	-	-	-0.145*	-0.270**	-
	angle (°)	0.365**	0.376**	0.186**	-0.401**	0.138**
Handle	weight (kg)	0.255**	-	0.147*	-	-
	wrist angle (°)	0.145*	-	-	-	-

* represents $p < 0.05$, ** represents $p < 0.01$.

5.3.4. The role of centrality measures

Degree, closeness, eigenvector, and betweenness centralities had different logics for calculating a node's importance. Figures 5.8 and 5.9 illustrate the different patterns of the cumulative percentage for UV, PA, and CV categories appearing in the top-n centrality measures.

The most prominent difference between the two datasets is the distribution of PA. Dataset 2 generally reveals a higher percentage of PA at higher ranks compared to Dataset 1. However, it is commonly observed that the set of PA includes the highest percentage in the top 10-betweenness centrality. This phenomenon indicates the role of PAs as mediators of other keywords. Therefore, it will be efficient to observe the betweenness centrality in the case of identifying PA from the text data.

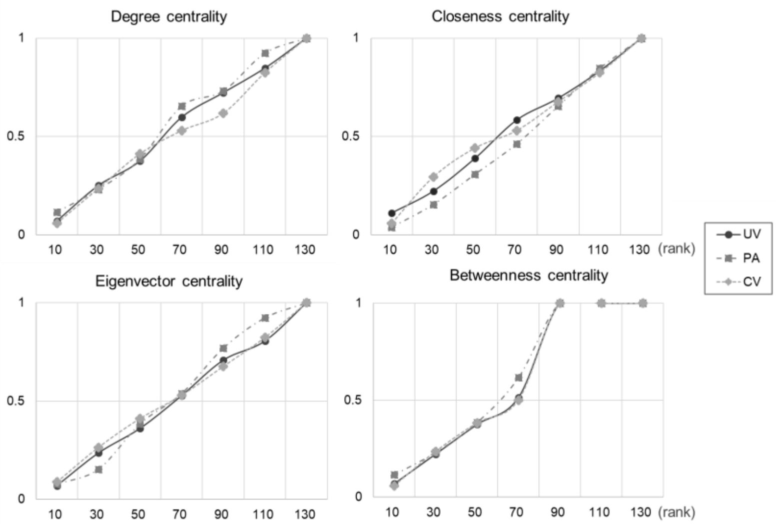


Figure 5.8. Cumulative percentage versus the ranking of UV, PA, and CV categories for Dataset 1

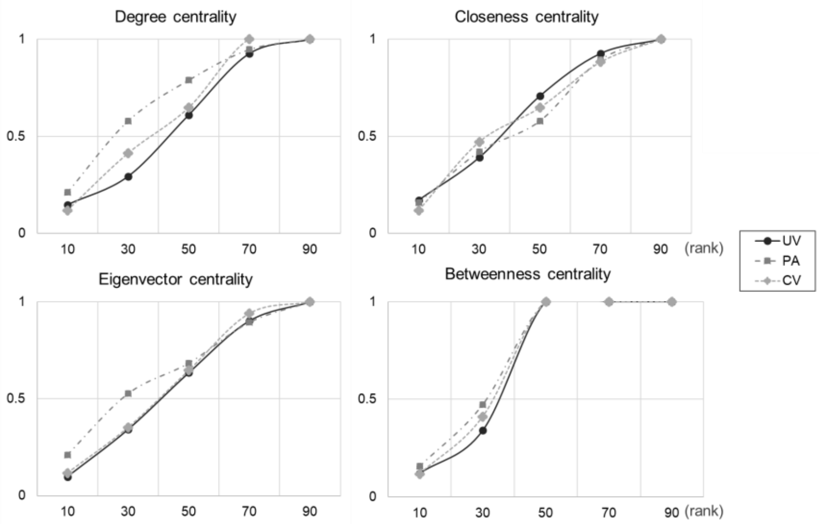


Figure 5.9. Cumulative percentage versus the ranking of UV, PA, and CV categories for Dataset 2

For degree centrality, there are relatively few UVs at high rank order (1–50th for Dataset 1, and 1–30th for Dataset 2). Meanwhile, eigenvector centrality shows a higher increasing rate of UV. This phenomenon indicates that high-rank UVs are related to small but important keywords. In the case of closeness centrality, a high number of UVs and CVs occur in top 10 and top 30, respectively. Considering nodes with high closeness centrality are independent from other nodes (Brass & Burkhardt, 1993), UV in the top 10 for closeness centrality are considered not confined to certain contexts or PA. Apart from UV in the top 10 keywords, CV shows higher closeness centrality than UV, revealing that most UVs were located in the local group of a network. In summary, it will be useful to examine different network centralities for different keyword categories; betweenness centrality for PA, closeness centrality and eigenvector centrality for UV, and closeness centrality for CV.

5.4. Discussion

This chapter presents a systematic approach to utilizing user-generated content in the form of unstructured text. After eliciting UV, a UX quantification model and related PAs were suggested based on qualitative research; the results were compared to the results of quantitative research, and the significance of the suggested method was demonstrated.

The first phase of the current study focused on collecting UV that influenced the overall satisfaction level. As a qualitative research, users' descriptions of vacuum cleaners under given contexts were collected and analyzed using a semantic network frame. Network representation allowed us to quantify the relationship between keywords, and keywords with high convergence in a network structure were considered to represent UV. In the second phase, network centralities were utilized as the weights of UV in UX quantification models. While building UX quantification models, seven network centralities (degree centrality, Bonacich power, Freeman betweenness centrality, flow betweenness centrality, eigenvector centrality, closeness centrality, and ARD) and term frequency were used to give weights to UV. The suggested models' goodness-of-fit were examined by calculating the R^2 and VIF values, and were compared to the results of the stepwise regression models. Finally, the semantic association between UV and PA were identified based on the assumption that the co-occurrence of UV and PA will represent their association. After selecting important PAs, PA-subnetworks were generated to examine the link strength and network centralities of UV. The relationships were then compared to the results of quantitative analysis. The result showed that the elicited PA were actually related to UV, but also revealed that the centrality measures were not proportional to the correlation coefficients.

From this study, we suggested a method of numerically presenting a UX based on qualitative textual data. Whereas existing studies have deliberately collected terms or suggested UX models on predefined hierarchical structures, this study took an inductive approach to exploring UV based on user expression data. We also suggested UX quantification models using centrality measures and term frequency as UV weights. Comparing R^2 values, term frequency showed lower values than centrality measures for both Datasets 1 and 2. However, the R^2 values between UX quantification models did not vary much, signifying the importance of selecting appropriate UVs. In this process, different network centralities can be applied to different keyword categories: betweenness centrality for PA, closeness centrality and eigenvector centrality for UV, and closeness centrality for CV.

In addition to UV, network centralities were also used to identify important PA. Semantic associations between UV and PA were observed by generating PA-subnetworks that showed significant relationships in the quantitative study. However, correlation analysis also showed significant relationships between variables that were not identified in a semantic network; for Dataset 1, “handle” and the UV of “following” and “quiet” were low but significant correlation coefficients, even though these terms did not co-occur. This phenomenon may have resulted from either simple coincidence or from some internal mechanisms of the products. Either way, researchers must pay more attention to interpreting such cases.

The suggested approach has several limitations. First, the effects of PA measurements in this paper are sufficient to understand the PAs of vacuum cleaners; besides the PAs that were measured in this study, additional metrics such as the center of mess or frictional force could have been used. Second, the values of the centrality measures were not proportional to the result of correlation analysis. The difference between these values can be interpreted as discrepancies between perceptual and the actual relatedness, as Patton (1990) described, "*Subjective data imply opinion rather than fact, intuition rather than logic, impression rather than confirmation.*" Despite these limitations, network analysis enabled the extraction of UV and PA when using vacuum cleaners, based on numerical representations.

The results of this study reveal the possibility of using semantic network analysis to quantify UX. Most centrality measures show slightly lower R^2 values than those of linear regression models. Considering that the coefficients of a linear regression model are inductively estimated by the maximum likelihood method, it is natural that UX quantification models have smaller R^2 values. However, UX researchers will be able to save much time and effort by using the suggested method rather than collecting numerical scores for each UV. For example, the method will help analyze the internet review data without the necessity of surveying quantitative scores. Considering that case studies cover issues of both usability and affective quality (look and feel), we expect that the suggested method can be adopted in various UX studies across devices and modalities.

CHAPTER 6

Conclusion and Discussion

6.1. Summary of findings

This dissertation has focused on developing a systematic research method of understanding UX based on qualitative text data. Chapter 1 presented the problems of qualitative research in understanding UX, and Chapter 2 reviewed literatures on semantic network analysis, UX evaluation technique, and product design. As noted in Chapter 1, none of the UX studies proposed a systematic research method that utilizes semantic network analysis on user expression data, despite its advantages of representing qualitative data with quantitative values. Therefore, a semantic network analysis method was mainly used throughout this dissertation. This dissertation tried to answer three Research Questions (RQ):

RQ 1: *Can we examine the representativeness of qualitative text data?*

RQ 2: *Can we extract UV and quantify their importance?*

RQ 3: *Can we identify relationships between UV and PA?*

The research objective was proposed to answer these research questions based on unstructured text data: examine representativeness (Chapter 3), elicit

UV that composes UX (Chapter 4), and quantify UX (Chapter 5)

In Chapter 3, the reliability of a qualitative text data was examined by adopting the concept of network stability. Among the semantic networks generated from text, subnetworks were sampled from the original networks until the representativeness of each sample size was determined. Then, similarities between the subnetworks and the original network were calculated by applying a correlation analysis to determine the network stability.

In Chapter 4, a network analysis was applied in eliciting UVs that affect users' satisfaction; top 10 keywords with the highest degree, closeness, betweenness, and eigenvector centralities were analyzed to elicit important UVs. Then, these UVs became independent variables of the consequent quantitative research; participants evaluated UVs on 7-point Likert scale. From this study, user satisfaction models for two user groups were suggested for a shutter press sound.

Chapter 5 proposed a method to build a UX quantification model by assuming that centrality measures can represent the effects of UV on user satisfaction. UX was expressed in a form of linear equation, by using network centralities as UV weights. The goodness-of-fit of the suggested model was verified via quantitative research. In addition, the relationship between UV and PA was investigated based on qualitative data. PA-subnetworks were generated, and UV's centralities in each subnetwork were calculated.

As summarized above, this dissertation gave positive answers to three Research Questions. Semantic network analysis enabled measuring the importance of UV, and predicting user's satisfaction based on qualitative research.

6.2. Practical implications of the research

This dissertation has shown how user expression data can be analyzed when using a semantic network analysis. A semantic network analysis enabled representing numerical values on concepts, which became UV in this dissertation. The node-level network centralities were mainly used in examining network stability, and quantifying the importance of UV and PA. This is an important step forward for UX study, as increasing number of textual data is collected by a technological development. Instead of interpreting the meaning of each sentence, UVs and PAs that compose UX could be suggested without involving a subjective judgement. The method also enabled researchers reduce time and effort, therefore it is expected to be practically used in developing a product.

First, the representativeness of user expression data was examined based on numerical inferences. Although this is not the first time that uses a network centrality to calculate the networks' similarities, to the best of our knowledge, no studies have reported the appropriateness of sample size of qualitative text data. For interview data which investigates relatively smaller number of participants, existing researchers subjectively determined if new information could be collected with larger populations (O'Reilly & Parker, 2013). Comparing to this approach, the proposed method has a competitive edge in terms of its required cost/effort and objectivity.

Second, important UVs and PAs were elicited based on network centralities; the centrality measures were used to give weights on UVs in building UX quantification model. The goodness-of-fit of these models was verified by comparing to the quantitative studies, and revealed the possibility of using network centrality as UV's importance. In addition, UX quantification models with centrality measures showed higher R^2 values, compared to the model that used term frequency as weight.

Third, relationships between UV and PA were identified based on the assumption that semantically associated terms will co-occur. During analysis, a subnetwork was generated for each PA, and centrality measures of UVs were calculated. Although the relationship strength between UV and PA should be analyzed by quantitative analysis (i.e. correlation analysis), the proposed method will help prevent the spurious correlation problem by examining the link strengths between keywords.

Throughout this dissertation, a systematic research method of using qualitative data was proposed. At data collection step, representativeness of text data was quantitatively measured by observing network stability. At analysis step, important UVs and PAs were elicited, and used as variables of quantitative research. Or, the elicited UVs were used to build UX quantification models by using network centralities as UV weights. At design step, relationships between UV and PA were identified. The result of qualitative text data assisted interpreting the result of quantitative analysis to prevent spurious correlation. Following these steps of data collection, analysis, and design will help understanding UX more effectively and efficiently.

6.3. Limitations and future research

There are limitations that should be noted. A semantic network analysis has its strength over the other methodologies as it automatically calculates the structural importance of concepts which reduces researchers' subjectivity, time, and efforts. However, it does not fall on the preprocessing step, as the process of defining synonyms and representative terms involves a subjective judgement. Especially, the process of categorizing keyword categories (UV, PA, and CV) and converting CV to UV still requires researchers to read and interpret each sentences. Whereas case studies in this dissertation involved four to five UX researchers to prevent subjective bias, further research is required to improve efficiency of the suggested method. Referring that several studies are trying to automate preprocessing step by applying linguistic resources such as Thesausus, the preprocessing step will also be automated in the future. Additionally, whereas the valence (positive, negative, neutral) of UVs were subjectively judged in this study, the process can be automated by applying sentiment analysis (Polanyi & Zaenen, 2006; Shaikh, Prendinger, & Mitsuru, 2007).

The question of generalizability means whether the suggested research method will also work for the other products and domains. In this dissertation, vacuum cleaners and camera shutter sounds were observed as case studies. The item encompassed visual, tactile, and auditory feelings related to the product's performance, but did not consider marketing or social factors such as brand image and customer loyalty. These additional factors could be investigated, if data were collected from the appropriate sources. In addition, it should be noted that we only observed relatively small number of participants. Although the method has its significance in analyzing large amounts of text data, we only have dealt with qualitative data collected from lab experiments, in order to compare the results to the quantitative research.

The last part of this dissertation deals with identifying semantic associations between UV and PA. However, the method is only useful for PA that general users can describe. Besides PAs which are externally visible and measurable, products also have internal PAs which are hard to be noticed without a domain knowledge (i.e. the size of a filter bag in a vacuum cleaner, shutter speed of a camera). Therefore, it is necessary to conduct a quantitative research for internal PA, as introduced in Chapter 4.

Lastly, performance of the suggested method was not compared to the existing qualitative research method, which reads and interprets sentences one by one. Although Van Atteveldt (2008) emphasized differences between semantic network and content analysis in abstraction level of concepts, it is still possible to generate networks with different abstraction levels, by defining terms as a higher concept. Since preprocessing step in this dissertation was not much different from traditional content analysis, the significance of the suggested approach was proved by conducting a quantitative research.

In further research, we hope the proposed method could be automatically processed, and help researchers understand UX more effectively and efficiently. Considering increasing amounts of user expression data are collected from Web or information systems, the suggested method will be useful for practical applications in understanding UX.

BIBLIOGRAPHY

Abilhoa, W. D., & De Castro, L. N. (2014). A keyword extraction method from twitter messages represented as graphs.*Applied Mathematics and Computation*,240, 308-325.

Adikari, S., McDonald, C., & Campbell, J. (2011). A design science framework for designing and assessing user experience *Human-computer interaction. design and development approaches* (pp. 25-34): Springer.

Albert, R., Jeong, H., & Barabási, A.-L. (1999). Internet: Diameter of the world-wide web. *nature, 401*(6749), 130-131.

Aldrich, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statistical science*, 364-376.

Altshuller, G. S. (1984). *Creativity as an exact science*: Gordon and Breach.

Antiqueira, L., Nunes, M. d. G. V., Oliveira Jr, O., & Costa, L. d. F. (2007). Strong correlations between text quality and complex networks features. *Physica A: Statistical Mechanics and its Applications, 373*, 811-820.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2007). *Show me the money!: deriving the pricing power of product features by mining consumer reviews.* Paper presented at the Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining.

Asche, D.-K. M., & Kreis, H. (2014). Apps as Crucial Value Components and their Impact on the Customer Experience. *Marketing Review St. Gallen, 31*(5), 42-51.

Attri, R., Grover, S., Dev, N., & Kumar, D. (2013). An ISM approach for modelling the enablers in the implementation of Total Productive Maintenance (TPM). *International Journal of System Assurance*

Engineering and Management, 4(4), 313-326.

Aures, W. (1985). Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale. *Acta Acustica united with Acustica, 59*(2), 130-141.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM, 8*, 361-362.

Baum, F. (2003). *The new public health*: Oxford University Press.

Bevan, N. (2001). International standards for HCI and usability. *International Journal of Human-Computer Studies, 55*(4), 533-552.

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing, 8*(4), 243-257.

Bisping, R. (1997). Car interior sound quality: Experimental analysis by synthesis. *Acta Acustica united with Acustica, 83*(5), 813-818.

Blauert, J., & Jekosch, U. (1997). Sound-quality evaluation–a multi-layered problem. *Acta Acustica united with Acustica, 83*(5), 747-753.

Bonacich, P. (1972). Technique for analyzing overlapping memberships. *Sociological methodology, 4*, 176-185.

Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory, 12*(1), 21-34.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). UCINET 6 for Windows. *Harvard: Analytic Technologies, 185*.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *science, 323*(5916), 892-895.

Bowen, D. L., & Carow, J. P. (2001). Use of consumer listening panels to enhance sound quality. *Appliance Engineer, May*, 1-7.

Boztepe, S. (2007). Toward a framework of product development for global markets: a user-value-based approach. *Design studies, 28*(5), 513-533.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37*(2), 204-215.

Bratko, I., & Šuc, D. (2003). Qualitative data mining and its applications. *CIT. Journal of computing and information technology, 11*(3), 145-150.

Brass, D. J., & Burkhardt, M. E. (1993). Potential power and power use: An investigation of structure and behavior.*Academy of management journal*, *36*(3), 441-470.

Bruseberg, A., & McDonagh-Philp, D. (2001). New product development by eliciting user experience and aspirations. *International Journal of Human-Computer Studies, 55*(4), 435-452.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods, 39*(3), 510-526.

Cabrera, D., Ferguson, S., & Schubert, E. (2007). 'Psysound3': Software for Acoustical and Psychoacoustical Analysis of Sound Recordings.

Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social science information, 22*(2), 191-235.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research, 42*(3), 294-320.

Carley, K. M. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological methodology, 23*(75-126).

Carley, K. M. (1997a). Extracting team mental models through textual analysis. *Journal of Organizational Behavior, 18*(s 1), 533-558.

Carley, K. M. (1997b). Network text analysis: The network position of concepts. *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts*, 79-100.

Carley, K. M., Columbus, D., & Azoulay, A. (2012). Automap user's guide 2012 (No. CMU-ISR-12-106). *Carnegie-mellon univ pittsburgh pa inst of software research international*.

Cha, Y.-S., Kim, K., Lee, J.-Y., Lee, J., Choi, M., Jeong, M.-H., … Oh, S.-R. (2011). MAHRU-M: A mobile humanoid robot platform based on a dual-network control system and coordinated task execution. *Robotics and Autonomous Systems, 59*(6), 354-366.

Chaffin, D. B., & Nelson, C. (2001). *Digital human modeling for vehicle and workplace design*: Society of Automotive Engineers Warrendale, PA.

Chang, C.-W., Lin, C.-T., & Wang, L.-Q. (2009). Mining the text information to optimizing the customer relationship management. *Expert Systems with applications, 36*(2), 1433-1443.

Chien, C.-F., Wang, H.-J., & Wang, M. (2007). A UNISON framework for analyzing alternative strategies of IC final testing for enhancing overall operational effectiveness. *International Journal of Production Economics, 107*(1), 20-30.

Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences, 3*(01), 1-15.

Coen, T., Jans, N., Van de Ponseele, P., Goethals, I., De Baerdemaeker, J., & De Moor, B. (2004). *Modelling the relationship between human perception and Sound Quality parameters using LS-SVMs.* Paper presented at the Proc. of the International Conference on Modal Analysis Noise and Vibration Engineering (ISMA 2004).

Cohen, L. (1988). Quality function deployment: An application perspective from digital equipment corporation. *Global Business and Organizational Excellence, 7*(3), 197-208.

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension.

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social networks, 25*(4), 283-307.

Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches*: Sage publications.

Crouch, M., & McKenzie, H. (2006). The logic of small samples in interview-based qualitative research. *Social science information, 45*(4), 483-499.

Csermely, P. (2006). *Weak links: Stabilizers of complex systems from proteins to social networks*: Springer.

Daniel, P., & Weber, R. (1997). Psychoacoustical roughness: Implementation of an optimized model. *Acta Acustica united with Acustica, 83*(1), 113-123.

Dedene, L., Valgaeren, R., Van Overmeire, M., & Guillaume, P. (1998). *Correlation of subjective sound perception of exhaust systems with sound quality metrics.* Paper presented at the Society for Experimental Mechanics, Inc, 16 th International Modal Analysis Conference.

Desmet, P. (2005). Measuring emotion: Development and application of an instrument to measure emotional responses to products. *Funology*, 111-123.

Diesner, J., & Carley, K. M. (2004). Revealing social structure from texts. *Causal mapping for research in information technology*, 81.

Diesner, J., & Carley, K. M. (2008). Conditional random fields for entity extraction and ontological text coding. *Computational and Mathematical Organization Theory, 14*(3), 248-262.

Doerfel, M. L. (1998). What constitutes semantic network analysis? A comparison of research and methodologies. *Connections*, *21*(2), 16-26.

Dorogovtsev, S. N., & Mendes, J. F. (2002). Evolution of networks. *Advances in physics*, *51*(4), 1079-1187.

Downe-Wamboldt, B. (1992). Content analysis: method, applications, and issues. *Health care for women international*, *13*(3), 313-321.

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). *Using latent semantic analysis to improve access to textual information.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological methods*, *5*(2), 155.

Elder, R. S., & Mohr, G. S. (2016). The crunch effect: Food sound salience as a consumption monitoring cue. *Food Quality and Preference, 51*, 39-46.

Estes, W. K. (1982). *Models of learning, memory, and choice: Selected papers*: Praeger New York.

Fairchild, M. D. (2013). Color appearance models. John Wiley & Sons, pp. 54.

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data, 2*(1), 5.

Fastl, H. (2000). *Sound quality of electric razors-Effects of Loudness.* Paper presented at the Proc. inter-noise.

Fastl, H. (2005). Recent developments in sound quality evaluation. *Proc. ForumAcusticum*, 1647-1653.

Fastl, H. (2006). *Psychoacoustic basis of sound quality evaluation and sound engineering.* Paper presented at the The Thirteenth International

Congress on Sound and Vibration.

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, 35*(3), 379-383.

Fechner, G. (1966). Elements of Psychophysics. Adler HE, translator; Howes DH, Boring EG, editors: New York: Holt, Rinehart & Winston.

Feng, J., Zhang, Y. Q., & Zhang, H. (2017). Improving the co-word analysis method based on semantic distance. *Scientometrics*, 1-11.

Flath, B., & Klein, E. (2014). *Advertising and Design: Interdisciplinary Perspectives on a Cultural Field*: transcript Verlag.

Foltz, P. W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, 28*(2), 197-202.

Franinovic, K., & Visell, Y. (2008). *Strategies for sonic interaction design: From context to basic design*.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks, 1*(3), 215-239.

Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social networks, 13*(2), 141-154.

Freeman, L. C. (2004). *The development of social network analysis*, Empirical Press Vancouver.

Ganglbauer, E., Schrammel, J., Deutsch, S., & Tscheligi, M. (2009). *Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility.* Paper presented at the Workshop on user experience evaluation methods in product development.

Glaser, B. G., Strauss, A. L., & Strutzel, E. (1968). The discovery of grounded theory; strategies for qualitative research. *Nursing research, 17*(4), 364.

Granovetter, M. (1976). Network sampling: Some first steps. *American Journal of Sociology, 81*(6), 1287.

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of consumer research, 5*(2), 103-123.

He, Y., & Tan, J. (2015). Study on SINA micro-blog personalized recommendation based on semantic network. *Expert Systems with Applications*,42(10), 4797-4804.

Hall, P., & Martin, M. A. (1988). On bootstrap resampling and iteration. *Biometrika*, *75*(4), 661-671.

Harway, N. I., & Iker, H. P. (1969). Content analysis and psychotherapy. *Psychotherapy: Theory, Research & Practice, 6*(2), 97.

Hassenzahl, M. (2005). The thing and I: understanding the relationship between user and product. *Funology*, 31-42.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität *Mensch & Computer 2003* (pp. 187-196): Springer.

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products–Facets of user experience. *Interacting with computers, 22*(5), 353-362.

Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology, 25*(2), 91-97.

Hauser, J. R., & Clausing, D. (1988). The house of quality. *Havard Business Review*, May-June 1988, 63-73.

Hoechstetter, M., Sautter, J.-M., Gabbert, U., & Verhey, J. (2016). Role of the Duration of Sharpness in the Perceived Quality of Impulsive Vehicle

Sounds. *Acta Acustica united with Acustica, 102*(1), 119-128.

Holbrook, M. B. (1999). *Consumer value: a framework for analysis and research*: Psychology Press.

Hong, Y. J., Shin, D., & Kim, J. H. (2016). High/low reputation companies' dialogic communication activities and semantic networks on Facebook: A comparative study. *Technological Forecasting and Social Change*, 110, 78-92.

Horvat, M., Domitrovi, H., & Jambrošić, K. (2012). Sound quality evaluation of hand-held power tools. *Acta Acustica united with Acustica, 98*(3), 487-504.

Hoser, B., Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006). *Semantic network analysis of ontologies*: Springer.

Hruschka, D. J., Schwartz, D., St. John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field methods, 16*(3), 307-331.

Huang, A. H., & Yen, D. C. (2013). Predicting the helpfulness of online reviews—a replication. *International Journal of Human-Computer Interaction, 29*(2), 129-138.

Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: the 10±2 rule. *Communications of the ACM, 53*(5), 130-133.

Iezzi, D. F. (2012). Centrality measures for text clustering. *Communications in Statistics-Theory and Methods*, 41(16-17), 3179-3197.

Iosif, E., & Potamianos, A. (2015). Similarity computation using semantic networks created from web-harvested data. *Natural Language Engineering, 21*(01), 49-79.

Iranmanesh, S. H., Thomson, V., & Salimi, M. H. (2005). Design parameter estimation using a modified QFD method to improve customer perception. *Concurrent Engineering, 13*(1), 57-67.

Ishimitsua, S., Sakamotob, K., Araic, T., Yoshimid, T., Fujimotod, Y., & Kawasakid, K. (2008). A study of evaluating the button sounds.

Isleifsdottir, J., & Larusdottir, M. (2008). *Measuring the user experience of a task oriented software.* Paper presented at the Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement.

ISO 226:223 (2003). Normal Equal-Loudness Level Contours. *International Organization for Standardization Geneva*.

ISO. 9241-210. (2009) Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. *International Standardization Organization (ISO). Switzerland*.

Jacobs, R. L., Leamer, E. E., & Ward, M. P. (1979). Difficulties with testing for causation. *Economic Inquiry, 17*(3), 401-413.

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., & Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. *Vaccine, 35*(29), 3621-3638.

Kelly, G. A. (1955). *The psychology of personal constructs. Volume 1: A theory of personality*: WW Norton and Company.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1/2), 81-93.

Kim, G. W., Lim, J., Choi, H., & Yun, M. H. (2012). *Adopting network analysis methods for contextual inquiry: The keyword structure representation of a web behavior.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Kim, G. W., Lim, J. H., & Yun, M. H. (2016). Analysis of Consumer Value Using Semantic Network: The Comparison of Hierarchical and Nonhierarchical Value Structures. *Human Factors and Ergonomics in Manufacturing & Service Industries, 26*(3), 393-407.

Kim, H. K., Han, S. H., Park, J., & Park, W. (2015). How user experience changes over time: A case study of social network services. *Human Factors and Ergonomics in Manufacturing & Service Industries, 25*(6), 659-673.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction, 22*(4-5), 441-504.

Kondracki, N. L., Wellman, N. S., & Amundson, D. R. (2002). Content analysis: Review of methods and their applications in nutrition education. *Journal of nutrition education and behavior, 34*(4), 224-230.

Kossinets, G. (2006). Effects of missing data in social networks. *Social networks, 28*(3), 247-268.

Krippendorff, K. (2004). Reliability in content analysis. *Human communication research, 30*(3), 411-433.

Kucera, H. (1985). *Language Design, Computers and Brains*. Retrieved from

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with computers, 23*(5), 473-483.

Kujala, S., & Väänänen-Vainio-Mattila, K. (2009). Value of information systems and products: Understanding the users' perspective and values. *Journal of Information Technology Theory and Application (JITTA), 9*(4), 4.

Kujala, S., Vogel, M., Pohlmeyer, A. E., & Obrist, M. (2013). *Lost in time: the* *meaning of temporal aspects in user experience.* Paper presented at the CHI'13 Extended Abstracts on Human Factors in Computing Systems.

Kulkarni, S. S., Apte, U. M., & Evangelopoulos, N. E. (2014). The use of latent semantic analysis in operations management research. *Decision Sciences, 45*(5), 971-994.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics, 22*(1), 79-86.

Kupper, L. L. (1978). *Applied regression analysis and other multivariate methods*: Duxbury Press.

Kurasaki, K. S. (2000). Intercoder reliability for validating conclusions drawn from open-ended interview data. *Field methods, 12*(3), 179-194.

Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in Human Behavior, 43*, 35-48.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes, 25*(2-3), 259-284.

Langeveld, L., Van Egmond, R., Jansen, R., & Özcan, E. (2013). Product sound design: Intentional and consequential sounds.

Law, E. L.-C., & Hvannberg, E. T. (2004). *Analysis of combinatorial user effect in international usability tests.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009). *Understanding, scoping and defining user experience: a survey approach.* Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems.

Law, E. L.-C., & van Schaik, P. (2010). Modelling user experience–An agenda for research and practice: Oxford University Press.

Law, E. L.-C., van Schaik, P., & Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies*, *72*(6), 526-541. doi:10.1016/j.ijhcs.2013.09.006

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of educational research, 52*(1), 31-60.

Lee, M., Rhie, Y. L., Kim, J., & Lim, J. H. (2014). *Development of a systematic process and automation tool for semantic network analysis on natural Language*. Paper presented at the the 15th APIEMS Conference, Jeju, Korea.

Lee, S.K. (2008). Objective evaluation of interior sound quality in passenger cars during acceleration. *Journal of Sound and Vibration, 310*(1), 149-168.

Lewis, J. R. (2001). Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. *International Journal of Human-Computer Interaction*, *13*(4), 445-479.

Lucene Korean Analyzer. (2012). Retrieved from http://cafe.naver.com/korlucene/.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203-208.

Lyon, R. H. (2003). Product sound quality-From perception to design. *Sound and vibration, 37*(3), 18-23.

Maguire, M. (2001). Methods to support human-centred design. *International Journal of Human-Computer Studies, 55*(4), 587-634.

Malandrakis, N., Potamianos, A., Iosif, E., & Narayanan, S. (2013). Distributional semantic models for affective text analysis. *IEEE Transactions on Audio, Speech, and Language Processing, 21*(11), 2379-2392.

Marshall, M. N. (1996). Sampling for qualitative research. *Family practice, 13*(6), 522-526.

Martin, M. K., Pfeffer, J., & Carley, K. M. (2013). Network text analysis of conceptual overlap in interviews, newspaper articles and keywords. *Social Network Analysis and Mining, 3*(4), 1165-1177.

McAuley, J., & Leskovec, J. (2013). *Hidden factors and hidden topics: understanding rating dimensions with review text.* Paper presented at the Proceedings of the 7th ACM conference on Recommender systems.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: A sourcebook. *Beverly Hills: Sage Publications*.

Moghimi, V., Jusan, M. B. M., Izadpanahi, P., & Mahdinejad, J. (2017). Incorporating user values into housing design through indirect user participation using MEC-QFD model. *Journal of Building Engineering, 9*, 76-83.

Moore, B. C., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society, 45*(4), 224-240.

Morgan, D. L. (1993). Qualitative content analysis: A guide to paths not taken. *Qualitative health research, 3*(1), 112.

Morse, J. M. (1994). Designing funded qualitative research. In Denzin, N & Lincoln, Y. (Eds.), Handbook of qualitative research (pp. 220–235) Thousand Oaks, CA, US: Sage Publications.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. *arXiv preprint arXiv:1306.5204*.

Müller, O., Schmiedel, T., Gorbacheva, E., & vom Brocke, J. (2016). Towards a typology of business process management professionals: identifying patterns of competences through latent semantic analysis. *Enterprise Information Systems, 10*(1), 50-80.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon. com.

Myint, S. (2003). A framework of an intelligent quality function deployment (IQFD) for discrete assembly environment. *Computers & Industrial Engineering, 45*(2), 269-283.

Nagamachi, M. (1995). Kansei engineering: a new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics, 15*(1), 3-11.

Nagamachi, M. (2002). Kansei engineering as a powerful consumer-oriented technology for product development. *Applied ergonomics, 33*(3), 289-294.

Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with applications, 36*(2), 2592-2602.

Ngulube, P. (2015). Qualitative data analysis and interpretation: systematic search for meaning. *Addressing Research Challenges: Making Headway for Developing Researchers. edn. Edited by Mathipa E, Gumbo M. Noordywk, South Africa: Mosala-MASEDI Publishers & Booksellers cc*, 131-156.

Nielsen, J. (1994). Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies, 41*(3), 385-397.

Nielsen, J., & Landauer, T. K. (1993). *A mathematical model of the finding of usability problems.* Paper presented at the Proceedings of the

INTERACT'93 and CHI'93 conference on Human factors in computing systems.

Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Nor, M. J. M., Fouladi, M. H., Nahvi, H., & Ariffin, A. K. (2008). Index for vehicle acoustical comfort inside a passenger car. *Applied Acoustics, 69*(4), 343-353.

Nykänen, A., & Sirkka, A. (2009). Specification of component sound quality applied to automobile power windows. *Applied Acoustics, 70*(6), 813-820.

Okamoto, S., Nagano, H., & Yamada, Y. (2013). Psychophysical dimensions of tactile perception of textures. *IEEE Transactions on Haptics, 6*(1), 81-93.

O'Reilly, M., & Parker, N. (2013). 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research, 13*(2), 190-197.

Oya, M. (2013). *Degree centralities, closeness centralities, and dependency distances of different genres of texts*. In Selected Papers from the 17th International Conference of Pan-Pacific Association of Applied Linguistics, 42-53.

Özcan, E., & Jacobs, J. (2014). Product sounds: Basic concepts and categories. *International Journal of Design, 8*(3), 97-111.

Özcan, E., & Schifferstein, H. N. (2014). *The effect of (un) pleasant sounds on the visual and overall pleasantness of products.* Paper presented at the Proceedings of the colors of care: The 9th international conference on design and emotion, 6-7 October 2014, Bogota, Colombia.

Pals, N., Steen, M. G., Langley, D. J., & Kort, J. (2008). Three approaches to take the user perspective into account during new product design. *International Journal of Innovation Management*, *12*(03), 275-294.

Pardo, T. A. S, Antiqueira, L., Nunes, M., Oliveira, O., & Costa, L. (2006). Modeling and evaluating summaries using complex networks. *Computational Processing of the Portuguese Language*, 1-10.

Park, J., & Han, S. H. (2013). Defining user value: A case study of a smartphone. *International Journal of Industrial Ergonomics, 43*(4), 274-282.

Park, J., Han, S. H., Kim, H. K., Moon, H., & Park, J. (2014). Developing and Verifying a Questionnaire for Evaluating User Value of a Mobile Device. *Human Factors and Ergonomics in Manufacturing & Service Industries*.

Park, J., Han, S. H., Kim, H. K., Oh, S., & Moon, H. (2013). Modeling user experience: A case study on a mobile device. *International Journal of Industrial Ergonomics, 43*(2), 187-196.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*: SAGE Publications, inc.

Patton, M. Q. (2005). *Qualitative research*: Wiley Online Library.

Pellegrini, R. S. (2001). *Quality assessment of auditory virtual environments*. Paper presented at the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001.

Petiot, J.-F., Kristensen, B. G., & Maier, A. M. (2013). *How should an electric vehicle sound? User and expert perception.* Paper presented at the ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

Pierson, R., & Bozmoski, A. (2003). Harley-Davidson's 100th Anniversary-The Sound of a Legend. *Sound and vibration, 37*(3), 14-17.

Pietila, G., & Lim, T. C. (2012). Intelligent systems approaches to product sound quality evaluations–A review. *Applied Acoustics, 73*(10), 987-1002.

Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters *Computing attitude and affect in text: Theory and applications* (pp. 1-10): Springer.

Popescu, A.-M., & Etzioni, O. (2007). Extracting product features and opinions from reviews *Natural language processing and text mining* (pp. 9-28): Springer.

Popping, R. (2003). Knowledge graphs and network text analysis. *Social science information, 42*(1), 91-106.

Quillian, M. R. (1969). The teachable language comprehender: a simulation program and theory of language. *Communications of the ACM, 12*(8), 459-476.

Raju, S., Shishtla, P., & Varma, V. (2009). *A Graph Clustering Approach to Product Attribute Extraction.* Paper presented at the IICAI.

Ramsey, C. A., & Hewitt, A. D. (2005). A methodology for assessing sample representativeness. *Environmental Forensics, 6*(1), 71-75.

Rayson, P., & Garside, R. (2000). *Comparing corpora using frequency profiling.* Paper presented at the Proceedings of the workshop on Comparing Corpora.

Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of advertising research, 28*(1), 11-31.

Rhiu, I., Ryu, T., Jin, B., & Yun, M. (2011). Analysis of customer satisfaction on the stiffness of outside panels of passenger cars. *Human-Computer Interaction. Users and Applications*, 257-265.

Rijsdijk, S. A., & Hultink, E. J. (2009). How today's consumers perceive tomorrow's smart products. *Journal of Product Innovation*

Management, 26(1), 24-42.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior, 12*(1), 1-20.

Rodríguez, M. A., & Egenhofer, M. J. (2003). Determining semantic similarity among entity classes from different ontologies. *IEEE transactions on knowledge and data engineering, 15*(2), 442-456.

Rokeach, M. (1968). A theory of organization and change within value-attitude systems. *Journal of Social Issues, 24*(1), 13-33.

Roto, V., Rantavuo, H., & Väänänen-Vainio-Mattila, K. (2009). *Evaluating user experience of early product concepts.* Paper presented at the Proc. DPPI.

Rousseau, R., & Zhang, Q. (1992). Zipf's data on the frequency of Chinese words revisited. *Scientometrics, 24*(2), 201-220.

Russell, J. A. (1978). Evidence of convergent validity on the dimensions of affect. *Journal of personality and social psychology, 36*(10), 1152.

Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. *European journal of operational research, 48*(1), 9-26.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information processing & management, 24(5), 513-523.

Savoy, A., Guo, Y., & Salvendy, G. (2009). Effects of importance and detectability of usability problems on sample size requirements. *Intl. Journal of Human–Computer Interaction, 25*(5), 430-440.

Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., & Jin, C. (2007). *Red Opal: product-feature scoring from reviews.* Paper presented at the Proceedings of the 8th ACM conference on Electronic commerce.

Schauerte, T. (2013). Identifying product attributes for quality function deployment: consumer perceptions in the case of wooden multi-storey houses. *Pro Ligno, 9*(4), 773-779.

Schmettow, M. (2012). Sample size in usability studies. *Communications of the ACM, 55*(4), 64-70.

Seth, G. (2008). Unstructured Data and the 80 Percent Rule: Investigating the 80%. *Clarabridge, Bridgepoints*, Q3.

Shaikh, M., Prendinger, H., & Mitsuru, I. (2007). Assessing sentiment of text by semantic dependency and contextual valence analysis. *Affective Computing and Intelligent Interaction*, 191-202.

Sharp, H., Rogers, Y., & Preece, J. (2007). Interaction design: beyond human-computer interaction, 339.

Shin, S.-H., Ih, J.-G., Hashimoto, T., & Hatano, S. (2009). Sound quality evaluation of the booming sensation for passenger cars. *Applied Acoustics, 70*(2), 309-320.

Shipley, B. (1999). Testing causal explanations in organismal biology: causation, correlation and structural equation modelling. *Oikos*, 374-382.

Sköld, A., Västfjäll, D., & Kleiner, M. (2004). *Predicting consumer preference from expert sensory ratings of sounds.* Paper presented at the Proceedings of the ICA 18th International Congress, Kyoto, Japan.

Smith, A., & Dunckley, L. (2002). Prototype evaluation and redesign: structuring the design space through contextual techniques. *Interacting with computers, 14*(6), 821-843.

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). *Recursive deep models for semantic compositionality over a sentiment treebank.* Paper presented at the Proceedings of the conference on empirical methods in natural language processing

(EMNLP).

Spence, C., & Shankar, M. U. (2010). The influence of auditory cues on the perception of, and responses to, food and drink. *Journal of Sensory Studies, 25*(3), 406-430.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science, 29*(1), 41-78.

Susini, P., McAdams, S., Winsberg, S., Perry, I., Vieillard, S., & Rodet, X. (2004). Characterizing the sound quality of air-conditioning noise. *Applied Acoustics, 65*(8), 763-790.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Sage.

Tanenbaum, W., & Brand, J. (2008). Using AutoMap for Social and Textual Network Analysis (No. ARL-TN-321). *Army research lab aberdeen proving ground md.*

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology, 42*(4), 253-264.

Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). *Feature-rich part-of-speech tagging with a cyclic dependency network.* Paper presented at the Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1.

Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery, 24*(3), 478-514.

Tuch, A. N., Bargas-Avila, J. A., Opwis, K., & Wilhelm, F. H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies, 67*(9), 703-715.

Tufte, E. (2003). The cognitive style of PowerPoint: pitching out corrupts within Cheshire: CT: Graphics Press.

Västfjäll, D., Gulbol, M.-A., Kleiner, M., & Gärling, T. (2002). Affective evaluations of and reactions to exterior and interior vehicle auditory quality. *Journal of Sound and Vibration, 255*(3), 501-518.

Västfjäll, D., Kleiner, M., & Gärling, T. (2003). Affective reactions to interior aircraft sounds. *Acta Acustica united with Acustica, 89*(4), 693-701.

Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections (Toronto, Ont.), 28*(1), 16.

Van Atteveldt, W. (2008). Semantic network analysis. *Techniques for Extracting, Representing, and Querying Media Content*.

Van Dantzig, S., Geleijnse, G., & van Halteren, A. T. (2013). Toward a persuasive mobile application to reduce sedentary behavior. *Personal and ubiquitous computing, 17*(6), 1237-1246.

Vermeeren, A. P., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). *User experience evaluation methods: current state and development needs.* Paper presented at the Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 34*(4), 457-468.

Visser, M., van Biljon, J., & Herselman, M. (2013). Evaluation of management information systems: A study at a further education and training college. *SA Journal of Information Management*, *15*(1), 8.

Wagner, V., & Kallus, K. W. (2013). *Use of a multidimensional approach in the automotive product development: Quality of turn indicator sounds.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference.

Wang, H., Chen, G., Lin, Z., & Wang, H. (2005). Algorithm of integrating QFD and TRIZ for the innovative design process. *International Journal of Computer Applications in Technology, 23*(1), 41-52.

Wang, X. (2015). Towards a methodology for consequential product sound design. *Mater's Thesis in the Master's Programme in Sound and Vibration, Gothenburg, Sweden*.

Warfield, J. N. (1973). On arranging elements of a hierarchy in graphic form. *IEEE Transactions on Systems, Man, and Cybernetics*(2), 121-132.

Warfield, J. N. (1978). Societal systems planning, policy and complexity. *Cybernetics and System*, *8*(1), 113-115.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8): Cambridge university press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’networks. *nature, 393*(6684), 440-442.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, *1*(6), 80-83.

Wilkins, A. J. (1971). Conjoint frequency, category size, and categorization time. *Journal of verbal learning and verbal behavior, 10*(4), 382-385.

Willett, J. B., & Keiley, M. K. (2000). Using covariance structure analysis to model change over time. *Handbook of applied multivariate statistics and mathematical modeling*, 665-694.

Woodruff, R. B. (1997). Customer value: the next source for competitive advantage. *Journal of the academy of marketing science, 25*(2), 139-153.

Wright, P. C., & Monk, A. F. (1991). A cost-effective evaluation method for use by designers. *International Journal of Man-Machine Studies, 35*(6), 891-912.

Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. *The Qualitative Report, 16*(3), 730.

Zampini, M., & Spence, C. (2004). The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *Journal of Sensory Studies, 19*(5), 347-363.

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *The Journal of marketing*, 2-22.

Zhang, L., Huang, S., Du, J., & Vertiz, A. (1996). *Identifying customer expectation of power window sound quality.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard studies in classical philology, 40*, 1-95.

Zwicker, E., & Fastl, H. (1999). Sharpness and sensory pleasantness *Psychoacoustics* (pp. 239-246): Springer.

Zwicker, E., & Fastl, H. (2013). *Psychoacoustics: Facts and models* (Vol. 22): Springer Science & Business Media.

APPENDIX A

Table A. Centrality measures of Shutter sound network

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
hard	0.851	1.000	0.117	1.000	0.307	0.859	0.918
sophisticated	0.776	0.978	0.058	0.638	0.300	0.807	0.881
tonality	0.746	0.953	0.056	0.514	0.293	0.788	0.866
good	0.746	0.949	0.057	0.482	0.292	0.788	0.866
silent	0.746	0.964	0.043	0.480	0.296	0.788	0.866
loud	0.701	0.911	0.048	0.390	0.281	0.761	0.843
annoying	0.687	0.912	0.045	0.467	0.281	0.753	0.836
classical	0.672	0.897	0.041	0.393	0.277	0.744	0.828
lingering	0.642	0.891	0.020	0.227	0.275	0.728	0.813
hollow	0.582	0.830	0.022	0.193	0.257	0.698	0.784
fun	0.582	0.806	0.031	0.336	0.251	0.698	0.784
lively	0.537	0.751	0.016	0.148	0.234	0.684	0.769
soft	0.537	0.694	0.036	0.362	0.218	0.677	0.761
friendly	0.507	0.726	0.012	0.115	0.227	0.663	0.746
balanced	0.507	0.728	0.016	0.168	0.228	0.663	0.746
ending	0.478	0.691	0.008	0.115	0.217	0.650	0.731
clean	0.478	0.708	0.008	0.096	0.222	0.650	0.731
charm	0.463	0.668	0.007	0.082	0.210	0.644	0.724
envelope shape	0.463	0.692	0.005	0.095	0.217	0.644	0.724
duration	0.463	0.688	0.005	0.076	0.216	0.644	0.724
pitch	0.463	0.693	0.005	0.058	0.218	0.644	0.724
overall	0.433	0.641	0.007	0.103	0.202	0.638	0.716
num of peaks	0.418	0.642	0.003	0.064	0.203	0.626	0.701
appropriate	0.418	0.619	0.008	0.146	0.196	0.632	0.709
comfortable	0.418	0.609	0.007	0.080	0.193	0.632	0.709
not good	0.388	0.574	0.005	0.058	0.183	0.615	0.687
rough	0.388	0.603	0.002	0.033	0.191	0.615	0.687

Table A. Centrality measures of Shutter sound network *(Continued)*

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
fresh	0.388	0.552	0.011	0.163	0.176	0.615	0.687
tempo	0.373	0.570	0.003	0.048	0.182	0.609	0.679
noise	0.373	0.521	0.009	0.136	0.167	0.609	0.679
loudness	0.358	0.545	0.003	0.055	0.175	0.609	0.679
discomfort	0.328	0.522	0.001	0.039	0.168	0.588	0.654
function	0.299	0.457	0.001	0.024	0.149	0.583	0.647
professional	0.299	0.462	0.001	0.022	0.151	0.578	0.639
metal	0.284	0.372	0.004	0.077	0.124	0.578	0.634
forepart	0.269	0.413	0.001	0.023	0.136	0.573	0.632
simple	0.239	0.365	0.001	0.023	0.122	0.568	0.619
percept	0.239	0.376	0.000	0.012	0.125	0.563	0.617
electronic	0.224	0.348	0.000	0.009	0.117	0.558	0.609
blur	0.224	0.298	0.001	0.027	0.102	0.549	0.604
film	0.224	0.313	0.003	0.096	0.107	0.554	0.602
positive	0.194	0.251	0.002	0.039	0.089	0.536	0.587
emotional	0.194	0.302	0.000	0.008	0.104	0.554	0.597
fricative	0.194	0.293	0.000	0.008	0.101	0.549	0.595
mirror	0.179	0.276	0.000	0.009	0.096	0.545	0.587
impact	0.179	0.267	0.000	0.011	0.094	0.545	0.587
complicated	0.179	0.266	0.000	0.018	0.093	0.545	0.587
material	0.164	0.243	0.000	0.006	0.086	0.540	0.580
vague	0.164	0.238	0.003	0.080	0.085	0.536	0.577
trust	0.149	0.217	0.000	0.007	0.079	0.536	0.572
viewfinder	0.134	0.149	0.000	0.030	0.059	0.523	0.555
impressive	0.134	0.197	0.000	0.004	0.073	0.532	0.565
feel taste	0.119	0.172	0.000	0.003	0.066	0.519	0.552
control	0.090	0.083	0.000	0.013	0.040	0.500	0.530
mirrorless	0.090	0.118	0.000	0.003	0.050	0.500	0.530
provocative	0.090	0.098	0.000	0.016	0.044	0.515	0.540
spring	0.075	0.093	0.000	0.001	0.043	0.504	0.527

Table A. Centrality measures of Shutter sound network *(Continued)*

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
motor sound	0.075	0.080	0.000	0.007	0.039	0.496	0.522
novel	0.060	0.064	0.000	0.001	0.034	0.500	0.520
plain	0.060	0.062	0.000	0.001	0.034	0.493	0.515
sticky	0.060	0.050	0.000	0.007	0.030	0.479	0.505
open close	0.045	0.045	0.000	0.000	0.029	0.496	0.512
button	0.045	0.016	0.000	0.009	0.020	0.447	0.475
work	0.045	0.027	0.000	0.002	0.024	0.462	0.488
consistent	0.045	0.024	0.000	0.001	0.023	0.450	0.478
stiff	0.045	0.034	0.000	0.000	0.026	0.479	0.500
expectation	0.030	0.000	0.000	0.001	0.016	0.469	0.488
delicate	0.030	0.012	0.000	0.000	0.019	0.482	0.498

APPENDIX B

Table B.1. Centrality measures of Pull and Push network

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
head	0.695	1.000	0.133	0.962	0.316	0.753	0.841
push feel	0.641	0.977	0.087	0.571	0.309	0.724	0.814
smooth	0.603	0.945	0.071	0.724	0.299	0.704	0.795
under furniture	0.580	0.924	0.065	0.515	0.293	0.693	0.784
light	0.519	0.874	0.049	0.628	0.277	0.665	0.753
power	0.496	0.882	0.031	0.391	0.279	0.652	0.740
noisy	0.450	0.753	0.047	0.514	0.239	0.636	0.719
heavy	0.450	0.790	0.034	0.412	0.250	0.636	0.719
handle	0.450	0.749	0.041	0.598	0.237	0.636	0.719
weak	0.427	0.762	0.025	0.285	0.241	0.624	0.706
maneuver	0.420	0.784	0.018	0.258	0.248	0.621	0.702
inconvenient	0.420	0.711	0.033	0.377	0.225	0.624	0.704
body	0.397	0.674	0.065	0.716	0.214	0.621	0.695
following	0.397	0.632	0.067	1.000	0.200	0.621	0.695
tight	0.366	0.684	0.018	0.263	0.217	0.601	0.676
convenient	0.328	0.644	0.011	0.200	0.204	0.590	0.658
stiff	0.298	0.613	0.007	0.151	0.194	0.577	0.641
flexible	0.290	0.600	0.006	0.113	0.190	0.577	0.641
rolling	0.290	0.558	0.009	0.192	0.177	0.575	0.637
fall	0.282	0.571	0.008	0.201	0.181	0.572	0.634
strong	0.282	0.617	0.004	0.111	0.196	0.572	0.634
push	0.282	0.566	0.008	0.165	0.180	0.575	0.635
thin	0.275	0.546	0.006	0.126	0.173	0.557	0.627
thick	0.267	0.542	0.005	0.108	0.172	0.550	0.618
freely	0.267	0.536	0.007	0.117	0.170	0.570	0.630
tough	0.260	0.544	0.008	0.255	0.173	0.567	0.623
dust	0.252	0.382	0.009	0.288	0.121	0.544	0.612
control	0.244	0.507	0.006	0.153	0.161	0.562	0.618
easy	0.244	0.492	0.010	0.200	0.156	0.560	0.615
at once	0.229	0.427	0.006	0.152	0.136	0.541	0.601
wheel	0.229	0.463	0.007	0.173	0.147	0.557	0.611

Table B.1. Centrality measures of Pull and Push network *(Continued)*

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
accessary	0.229	0.395	0.005	0.184	0.125	0.528	0.593
power control	0.221	0.469	0.004	0.130	0.149	0.541	0.598
several times	0.221	0.394	0.004	0.158	0.125	0.539	0.597
movement	0.214	0.479	0.002	0.118	0.152	0.550	0.599
good	0.206	0.439	0.004	0.120	0.139	0.546	0.594
body weight	0.198	0.332	0.027	0.368	0.105	0.555	0.599
hose	0.198	0.407	0.003	0.128	0.129	0.546	0.594
floor	0.198	0.455	0.002	0.104	0.144	0.535	0.587
roller	0.183	0.416	0.001	0.067	0.132	0.533	0.583
deep	0.183	0.438	0.001	0.028	0.139	0.544	0.588
pull	0.183	0.401	0.002	0.075	0.127	0.530	0.581
necessary	0.176	0.380	0.001	0.051	0.121	0.522	0.571
grip	0.168	0.381	0.001	0.030	0.121	0.512	0.565
appropriate	0.160	0.322	0.003	0.136	0.102	0.533	0.571
stick	0.153	0.312	0.003	0.194	0.099	0.533	0.569
intrusive	0.145	0.264	0.001	0.079	0.084	0.500	0.550
get in	0.145	0.365	0.000	0.023	0.116	0.530	0.565
lay	0.137	0.294	0.001	0.051	0.094	0.498	0.546
big	0.130	0.296	0.002	0.072	0.094	0.524	0.559
annoying	0.130	0.254	0.001	0.060	0.081	0.508	0.547
vacuum	0.122	0.211	0.005	0.197	0.067	0.524	0.553
head weight	0.122	0.251	0.000	0.044	0.080	0.506	0.546
fatigue	0.122	0.229	0.001	0.106	0.073	0.508	0.543
control feeling	0.122	0.292	0.000	0.032	0.093	0.494	0.536
small	0.115	0.242	0.002	0.078	0.077	0.508	0.545
burdensome	0.115	0.200	0.002	0.118	0.064	0.502	0.537
overall	0.115	0.280	0.001	0.040	0.089	0.522	0.550
wrist	0.115	0.241	0.000	0.042	0.077	0.496	0.537
pop out	0.115	0.234	0.001	0.088	0.074	0.492	0.534
speed	0.115	0.217	0.000	0.037	0.069	0.489	0.532

Table B.1. Centrality measures of Pull and Push network *(Continued)*

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
fast	0.115	0.217	0.000	0.037	0.069	0.489	0.532
urgent	0.115	0.217	0.000	0.037	0.069	0.489	0.532
height	0.115	0.217	0.000	0.037	0.069	0.489	0.532
handle weight	0.107	0.230	0.000	0.028	0.073	0.510	0.542
impossible	0.107	0.203	0.000	0.094	0.065	0.482	0.522
angle	0.107	0.203	0.000	0.053	0.065	0.496	0.533
corner	0.107	0.263	0.001	0.044	0.084	0.524	0.551
center	0.107	0.258	0.000	0.009	0.082	0.485	0.527
scratch	0.099	0.208	0.000	0.051	0.066	0.504	0.536
bumper	0.099	0.208	0.000	0.051	0.066	0.504	0.536
assistant wheel	0.099	0.198	0.000	0.043	0.063	0.482	0.522
malfun-ction	0.099	0.198	0.000	0.043	0.063	0.482	0.522
frequent	0.099	0.198	0.000	0.043	0.063	0.482	0.522
plain	0.099	0.198	0.000	0.043	0.063	0.482	0.522
rough	0.099	0.232	0.000	0.028	0.074	0.489	0.527
separate	0.099	0.232	0.000	0.028	0.074	0.489	0.527
adjust	0.099	0.232	0.000	0.028	0.074	0.489	0.527
usability	0.092	0.184	0.000	0.037	0.059	0.491	0.522
lift	0.092	0.144	0.004	0.167	0.046	0.471	0.508
discord	0.092	0.216	0.000	0.043	0.069	0.500	0.529
coming and going	0.092	0.170	0.000	0.090	0.054	0.485	0.519
bend	0.084	0.152	0.001	0.076	0.049	0.487	0.520
neighbor	0.084	0.210	0.000	0.017	0.067	0.500	0.529
microdust	0.084	0.221	0.000	0.005	0.071	0.489	0.522
elimination	0.084	0.165	0.000	0.040	0.053	0.485	0.519
falling	0.084	0.165	0.000	0.040	0.053	0.485	0.519
stick weight	0.084	0.182	0.000	0.025	0.058	0.483	0.518
bow	0.076	0.203	0.000	0.006	0.065	0.482	0.514
rubber	0.076	0.162	0.000	0.030	0.052	0.485	0.517
prevent	0.076	0.162	0.000	0.030	0.052	0.485	0.517

Table B.1. Centrality measures of Pull and Push network *(Continued)*

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
pile up	0.076	0.167	0.000	0.056	0.053	0.478	0.509
width	0.069	0.158	0.000	0.037	0.050	0.487	0.515
high place	0.069	0.163	0.000	0.025	0.052	0.468	0.501
many	0.069	0.163	0.000	0.025	0.052	0.468	0.501
attach	0.069	0.149	0.000	0.024	0.048	0.475	0.506
position	0.069	0.149	0.000	0.024	0.048	0.475	0.506
head gap	0.069	0.163	0.000	0.023	0.052	0.478	0.509
shoulder	0.061	0.099	0.000	0.064	0.032	0.444	0.480
arm	0.061	0.099	0.000	0.064	0.032	0.444	0.480
auto	0.061	0.099	0.000	0.064	0.032	0.444	0.480
rotation	0.061	0.178	0.000	0.001	0.057	0.508	0.525
connection part	0.061	0.140	0.000	0.026	0.045	0.492	0.515
damage	0.061	0.140	0.000	0.026	0.045	0.492	0.515
spread	0.053	0.067	0.000	0.090	0.022	0.443	0.473
long	0.053	0.119	0.000	0.030	0.038	0.478	0.501
power button	0.053	0.096	0.000	0.014	0.031	0.450	0.481
again	0.053	0.115	0.000	0.008	0.037	0.482	0.506
evenly	0.046	0.065	0.000	0.043	0.021	0.443	0.473
design	0.046	0.119	0.000	0.001	0.038	0.455	0.482
stable	0.046	0.119	0.000	0.002	0.038	0.471	0.495
usual	0.046	0.047	0.030	0.503	0.015	0.409	0.439
not good	0.046	0.101	0.000	0.008	0.033	0.466	0.491
worried	0.038	0.061	0.000	0.098	0.020	0.423	0.449
body size	0.038	0.075	0.000	0.014	0.024	0.431	0.459
waist	0.038	0.068	0.000	0.007	0.022	0.421	0.450
crude	0.038	0.053	0.000	0.194	0.017	0.456	0.480
cornering	0.038	0.077	0.000	0.009	0.025	0.434	0.462
side	0.038	0.084	0.000	0.009	0.027	0.438	0.467
possible	0.038	0.080	0.000	0.009	0.026	0.441	0.467
sweep	0.038	0.082	0.000	0.003	0.026	0.425	0.454

Table B.1. Centrality measures of Pull and Push network *(Continued)*

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
neighbor noise	0.031	0.059	0.000	0.056	0.019	0.423	0.449
unstable	0.031	0.051	0.000	0.103	0.017	0.456	0.480
apart	0.031	0.051	0.000	0.103	0.017	0.456	0.480
carpet	0.031	0.079	0.000	0.001	0.025	0.461	0.482
cord	0.031	0.056	0.000	0.008	0.018	0.415	0.441
alike	0.023	0.013	0.000	0.015	0.004	0.364	0.380
tick over	0.023	0.051	0.000	0.001	0.017	0.425	0.449
left and right	0.023	0.061	0.000	0.000	0.020	0.428	0.452
without problem	0.015	0.036	0.000	0.000	0.012	0.396	0.417
familiar	0.015	0.000	0.000	0.084	0.000	0.292	0.307
usable	0.015	0.000	0.000	0.084	0.000	0.292	0.307

Table B.2. Centrality measures of Storage network

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
storage	0.684	1.000	0.204	1.000	0.428	0.752	0.836
inconvenient	0.553	0.816	0.144	0.886	0.351	0.673	0.765
spacious	0.526	0.872	0.072	0.621	0.375	0.673	0.757
handle	0.474	0.647	0.133	0.459	0.279	0.650	0.730
body	0.474	0.815	0.058	0.276	0.351	0.650	0.730
vertical	0.447	0.744	0.078	0.680	0.321	0.633	0.715
stick fixation	0.421	0.732	0.065	0.543	0.316	0.628	0.704
hose	0.368	0.687	0.034	0.232	0.297	0.613	0.684
unstable	0.355	0.587	0.040	0.392	0.255	0.598	0.669
light	0.303	0.621	0.009	0.171	0.269	0.576	0.640
stable	0.303	0.505	0.057	0.446	0.220	0.576	0.640
stick	0.303	0.593	0.024	0.255	0.258	0.555	0.629
head	0.289	0.586	0.010	0.100	0.255	0.576	0.640
small	0.263	0.564	0.004	0.069	0.246	0.543	0.610
volume	0.237	0.464	0.007	0.155	0.204	0.531	0.599
separation	0.237	0.360	0.016	0.172	0.159	0.547	0.603
convenient	0.237	0.519	0.004	0.035	0.227	0.555	0.607
heavy	0.224	0.477	0.004	0.047	0.209	0.543	0.601
droop	0.197	0.432	0.004	0.056	0.190	0.547	0.594
shape	0.197	0.385	0.010	0.132	0.170	0.543	0.592
design	0.184	0.365	0.029	0.338	0.162	0.543	0.588
body size	0.184	0.370	0.005	0.085	0.164	0.524	0.577
big	0.171	0.280	0.006	0.156	0.126	0.507	0.557
joint	0.171	0.366	0.005	0.071	0.162	0.543	0.583
stick length	0.171	0.227	0.015	0.267	0.104	0.510	0.559
move	0.158	0.184	0.013	0.247	0.086	0.524	0.568
fall	0.158	0.340	0.001	0.060	0.151	0.494	0.548
color	0.158	0.348	0.000	0.026	0.155	0.503	0.555
angulate	0.158	0.348	0.000	0.026	0.155	0.503	0.555
body weight	0.158	0.348	0.000	0.026	0.155	0.503	0.555
position	0.145	0.134	0.004	0.211	0.064	0.481	0.531

Table B.2. Centrality measures of Storage network *(Continued)*

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
stiff	0.145	0.298	0.001	0.042	0.134	0.521	0.557
long	0.145	0.349	0.000	0.012	0.155	0.531	0.568
durability	0.132	0.176	0.005	0.089	0.082	0.500	0.539
in the air	0.132	0.296	0.000	0.045	0.133	0.481	0.526
tidy	0.118	0.237	0.003	0.112	0.108	0.510	0.546
at once	0.118	0.282	0.000	0.018	0.127	0.481	0.526
break	0.105	0.143	0.002	0.101	0.068	0.490	0.524
warm	0.105	0.266	0.000	0.006	0.120	0.494	0.531
balanced	0.105	0.217	0.000	0.019	0.100	0.487	0.526
luxury	0.105	0.217	0.000	0.019	0.100	0.487	0.526
loose	0.105	0.089	0.000	0.172	0.045	0.425	0.471
old fashion	0.105	0.162	0.000	0.066	0.076	0.472	0.515
lay	0.105	0.162	0.000	0.066	0.076	0.472	0.515
recent	0.105	0.162	0.000	0.066	0.076	0.472	0.515
handle position	0.105	0.162	0.000	0.066	0.076	0.472	0.515
arrange	0.092	0.213	0.000	0.021	0.098	0.463	0.504
length	0.092	0.213	0.000	0.021	0.098	0.463	0.504
rubber	0.092	0.083	0.000	0.078	0.043	0.425	0.471
plastic	0.092	0.083	0.000	0.078	0.043	0.425	0.471
attach detach	0.092	0.083	0.000	0.078	0.043	0.425	0.471
dustbin	0.079	0.076	0.000	0.091	0.040	0.455	0.493
not friendly	0.079	0.076	0.000	0.091	0.040	0.455	0.493
waist	0.079	0.076	0.000	0.091	0.040	0.455	0.493
bow	0.079	0.076	0.000	0.091	0.040	0.455	0.493
kids	0.079	0.159	0.000	0.011	0.075	0.461	0.498
unusual	0.079	0.190	0.000	0.006	0.088	0.490	0.520
slanted	0.079	0.216	0.000	0.001	0.099	0.503	0.529
noisy	0.079	0.087	0.000	0.100	0.045	0.463	0.496
lifting	0.066	0.069	0.000	0.027	0.037	0.422	0.461
low	0.066	0.084	0.000	0.022	0.044	0.418	0.452

Table B.2. Centrality measures of Storage network *(Continued)*

	Degree	BonPwr	Betwee	FlowBet	Eigenv	Closen	ARD
hands	0.066	0.082	0.000	0.045	0.043	0.463	0.496
usual	0.053	0.123	0.000	0.047	0.060	0.475	0.500
terrace	0.053	0.123	0.000	0.047	0.060	0.475	0.500
weight	0.053	0.119	0.000	0.048	0.059	0.455	0.485
appropriate	0.053	0.119	0.000	0.048	0.059	0.455	0.485
good	0.053	0.123	0.000	0.003	0.060	0.455	0.485
usable	0.053	0.072	0.000	0.017	0.039	0.439	0.467
burdensome	0.053	0.053	0.000	0.019	0.031	0.400	0.430
soon	0.039	0.098	0.000	0.001	0.050	0.461	0.485
intuitive	0.039	0.025	0.000	0.013	0.019	0.380	0.407
wheel	0.039	0.044	0.000	0.068	0.027	0.450	0.476
rotation	0.039	0.044	0.000	0.068	0.027	0.450	0.476
crude	0.026	0.000	0.000	0.000	0.008	0.353	0.368
difficult	0.026	0.019	0.000	0.000	0.016	0.388	0.408
cool	0.026	0.020	0.000	0.000	0.017	0.390	0.411
round	0.013	0.006	0.000	0.000	0.011	0.367	0.386

국 문 초 록

사용자 경험(UX; User Experience)을 이해하는 데 있어 정성적 연구는 필수적이다. 정량적 연구가 통계적 분석을 통해 현상을 일반화하는 반면, 정성적 연구는 제품이나 서비스의 사용 맥락, 느낌, 태도 등 언어적 표현을 기반으로 현상을 이해하는 것을 목적으로 한다. 이러한 특성으로 인해 정성적 연구 방법은 수집한 데이터를 읽고, 의미를 이해하고, 요약하는 과정에서 시간과 노력이 들 뿐 아니라 해석 시 연구자의 주관적 의견이 개입할 수 있다. 따라서 본 논문은 의미망 분석(semantic network analysis) 을 활용하여 정성적 데이터를 체계적으로 분석하는 방법론을 제시하였다. 의미망 분석은 언어적인 표현을 네트워크의 구조로 표상함으로써 표현의 주체가 가지고 있는 인식과 개념의 연관 관계를 추정하는 방법론이다.

본 논문은 사용자 경험 연구 흐름에 따른 3가지 주제를 제시하고, 연구자들의 주관적인 의견을 최소화할 수 있는 체계적인 분석 방법론을 제시하였다. 각 주제는 다음과 같다: (1) 데이터 수집 과정에서 정성적 데이터의 대표성을 확보한다. (2) UX를 구성하는 주요 사용자 가치(UV; User Value) 를 추출한다. (3) 네트워크의 정량적인 수치를 활용하여 UX 모형을 제시하고 제품 요소(PA; Product Attribute)와의 관계를 설명한다.

첫 번째로, 데이터의 대표성을 검정하기 위해 의미망이 안정화되는 정도를 측정하였다. 초기 소수의 텍스트를 기반으로 구성된 의미망은 안정적이지 않고, 새로운 텍스트가 추가될 때마다 네트워크의 구조에 큰 변동이 생긴다. 그러나 의미망이 충분히 안정화되면, 유사한 데이터의 추가에도 그 구조가 크게 변화하지 않을 것으로 예측할 수 있다. 본 논문에서는 두 개의 정성적 인터뷰 데이터와 하나의 인터넷 리뷰 데이터를 대상으로 네트워크 안정성을 관찰한 결과, 본 연구 방법론이 소수 데이터뿐 아니라, 대량의 데이터에도 적용 가능하다는 것을 볼 수 있었다.

두 번째로, 연구자의 주관적 의견을 최소화하여 주요 UV를 추출하였다. 자연어를 수집한 후, 전처리 과정을 거친 뒤 네트워크를 형성하면 각 노드는 하나의 컨셉을 대표하고, 링크는 컨셉과 컨셉 사이의 연결 관계를 나타내게 된다. 네트워크의 구조적 중요도를 표상하는 네 가지 유형의 중심성(degree centrality, closeness centrality, betweenness centrality, eigenvector centrality) 에서 상위 10위에 드는 키워드를 기반으로 UV를 정의하였으며, 이어지는 정량적 연구에서 각 UV에 대한 점수를 조사함으로써 사용자 만족도 모형을 제시하였다.

세 번째로, 선형 방정식의 형태로 UX를 표현하였다. 각 노드의 중심성이 UV의 중요도를 나타낸다는 가정 하에, 일곱 가지의 중심성 척도가 각 UV의 가중치로 활용되었다. 본 연구에서는 세 개의 인터뷰 데이터를 토대로 UX 정량화 모형을 제안하였으며, 모형의 적합도를 검정하기 위해 결정계수와 (R^2 값) 다중공선성 수치(VIF)를 계산하였다. 또한, PA와 UV의 관계를 의미적 수준에서 관찰하였다.

제품의 물리적 특성 과 UV의 상관관계는 쉽게 관찰될 수 있으나, 통계적 분석 결과가 우연에 의한 것인지, 인과관계에 의한 것인 지 설명하기 위해서는 추가적인 이해가 수반되어야 한다. 따라서, 본 논문에서는 UV를 표상하는 단어와의 공동 출현 (co-occurrence)을 기반으로 의미 수준에서의 연관성을 도출하였다.

본 연구에서 제시된 방법론은 정성적 데이터를 분석하는 데 있어 연구자의 주관적인 의견 개입을 최소화하고 분석 시간과 노력을 줄일 수 있다는 데 의의가 있다. 데이터 신뢰성을 확보하기 위해서는 네트워크 안정성을 측정하는 방법론을 제시하였는데, 본 방법론을 적용한다면 데이터 수집 시 비용과 시간을 최소화할 수 있을 것이다. 정성적 데이터를 분석하는 과정에는 네트워크 중심성 값을 활용하여 UX 정량화 모형을 제시하였다. 이와 같이 정량적 데이터를 수치로 표상함으로써, UV를 추출한 후 각 변수에 대한 정량적 데이터를 수집하는 단계를 줄일 수 있었다. 마지막으로, 사용자들이 인식하고 있는 UV와 PA의 관계를 파악하였다. 이 단계에서 사용자 표현 데이터는 정량적 연구 결과가 우연에 의한 것인지, 인과관계에 의한 것인지 판단하는 보조적인 수단으로 사용되었다.

본 논문은 의미망 분석을 사용하여 사용자 표현 데이터를 분석하는 방법을 제안하였다. 최근 정보통신 기술 발전에 따라 대량의 비정형적 데이터 수집이 용이해지면서, 보다 빠르고 실용적인 분석 방법이 개발될 필요가 늘고 있다. 이 논문에서 소개한 방법론은 기존의 정성적 데이터를 이해하고 해석하는 과정을 단축할 수 있으므로, 더 효율적으로 UX를 이해할 수 있을 것이다.

주요어: 사용자 경험, 정성적 연구, 텍스트 데이터, 의미망 분석, 네트워크 안정성, 데이터 대표성, UX 정량화 모형, 제품 요소

학 번: 2012-21066